Cogn Process (2013) 14:129–142
DOI 10.1007/s10339-013-0539-3

RESEARCH REPORT

When do spatial abilities support student comprehension of STEM visualizations?

Scott R. Hinze · Vickie M. Williamson · Mary Jane Shultz · Kenneth C. Williamson · Ghislain Deslongchamps · David N. Rapp

Received: 28 November 2012 / Accepted: 16 January 2013 / Published online: 5 February 2013

Abstract Spatial visualization abilities are positively related to performance on science, technology, engineering, and math tasks, but this relationship is influenced by task demands and learner strategies. In two studies, we illustrate these interactions by demonstrating situations in which greater spatial ability leads to problematic performance. In Study 1, chemistry students observed and explained sets of simultaneously presented displays depicting chemical phenomena at macroscopic and particulate levels of representation. Prior to viewing, the students were asked to make predictions at the macroscopic level. Eye movement analyses revealed that greater spatial ability was associated with greater focus on the prediction-relevant macroscopic level. Unfortunately, that restricted focus was also associated with lower-quality explanations of the phenomena. In Study 2, we presented the same displays but manipulated whether participants were asked to make predictions prior to viewing. Spatial ability was again associated with restricted focus, but only for students who completed the prediction task. Eliminating the prediction task encouraged attempts to integrate the displays that related positively to performance, especially for participants with high spatial ability. Spatial abilities can be recruited in effective or ineffective ways depending on alignments between the demands of a task and the approaches individuals adopt for completing that task.

Keywords Spatial cognition · STEM · Visualizations · Learning

This article is part of the special issue on "Spatial Learning and Reasoning Processes," guest-edited by Thomas F. Shipley, Dedre Gentner.

Handling editor: Thomas F. Shipley.

S. R. Hinze · D. N. Rapp (✉)
Department of Psychology and School of Education & Social Policy, Northwestern University, 2120 Campus Drive, Evanston, IL 60208, USA
e-mail: rapp@northwestern.edu

V. M. Williamson
Department of Chemistry, Texas A&M University, College Station, TX, USA

M. J. Shultz
Department of Chemistry, Tufts University, Medford, TX, USA

K. C. Williamson
Department of Construction Science, Texas A&M University, College Station, TX, USA

G. Deslongchamps
Department of Chemistry, University of New Brunswick, Fredericton, Canada

Students in science, technology, engineering, and math (STEM) fields must think about processes and relationships at spatial or temporal scales that are unobservable to the naked eye. Chemistry students may be asked to understand how the movement of submicroscopic particles underlies the solid, liquid, and gas phases of water (e.g., Johnstone 1993). Earth science students need to infer how structural layers of land masses have developed over millions of years (e.g., Kali and Orion 1996). Physics students are required to visualize and predict the motion of imagined objects in three-dimensional space (Kozhevnikov et al. 2007). Based on the spatially complex requirements of many activities within STEM domains, it is not surprising that individual differences in spatial skills are positively related to performance in STEM courses (e.g., Rochford 1985; Wu and Shah 2004) and on STEM laboratory tasks

(Höffler 2010). Spatial skills are also positively related to long-term professional involvement with and success in STEM fields (Shea et al. 2001; Wai et al. 2009). But are there situations in which the application of spatial skills might create challenges for performance?

While positive correlations between spatial skills and STEM outcomes have been consistently documented in the extant literature, the relationship between spatial thinking and STEM learning is potentially complicated and less direct than might be expected (Uttal and Cohen 2012). Evidence for this complexity has emerged from two lines of research, providing clues as to when spatial reasoning will or will not be effectively applied. First, spatial thinking involves a variety of different types of skills and abilities, making it crucial to consider the match between the particular spatial skills of interest and the demands of the target task to which these skills may be deployed (Newcombe and Shipley in press). Second, beyond possessing particular sets of spatial skills, learners develop diverse strategies for when and how they might apply these skills to complete tasks (see Hegarty 2010).

To date, the types of spatial skills and the situations under which they might be effectively applied have received considerable attention. A variety of taxonomies have been proposed to summarize these interactions by outlining how spatial experiences invoke particular cognitive processes (Carroll 1993; Hegarty and Waller 2005; Linn and Peterson 1985; Pellegrino and Kail 1982; Thurstone 1938). Many of these taxonomies distinguish between particular skills by outlining the different types of activities that might be accomplished with them. For example, *spatial visualization* involves the ability to mentally represent (i.e., to imagine) and dynamically manipulate objects. This skill is particularly useful in STEM settings, such as when students and practitioners attempt to infer or interpret two-dimensional cross-sections of larger objects including layers of the Earth, geometric figures, and representations of the human body (Keehner et al. 2008). Other taxonomies differentiate spatial skills based on both the *type* of information being presented and the nature of the *task* that guides interaction with that information (Newcombe and Shipley in press). For example, presented information can be categorized as intrinsic (i.e., specific to a particular object including its form and orientation) or extrinsic (i.e., based on the relationship between the object and the real world), and tasks can be differentiated as static (involving fixed objects) or dynamic (involving moving or movable objects). These taxonomic descriptions prove useful for delineating when and how particular skills might be applied to accomplish particular goals and for determining conditions for which spatial thinking might be less than effective. The goal of the current project was not to test the validity of these taxonomies, but to consider situations derived from them for which the demands of a target task might misalign with the strategic allocation of spatial skills. It is precisely these conditions that should reveal reduced benefits of spatial skills on STEM performance, in contrast to the more generalized benefits consistently reported in the literature.

Consider the following examples to illustrate these situations. In order to put together a child's bike from illustrated instructions, it is necessary to visualize and manipulate the images from the instructions to match the size and orientation of actual bike parts (e.g., Brunyé et al. 2006). An intrinsic-static skill, such as identifying embedded figures, would prove less relevant to accomplishing the task than an intrinsic-dynamic skill, such as mental rotation (Vandenberg and Kuse 1978). In STEM classrooms, an intrinsic-dynamic spatial skill such as mental rotation, however, would not be particularly useful for memorizing the features of a topographical map (e.g., Sanchez and Branaghan 2008) or identifying chemical structure from a static image of a molecule, as these activities may rely more on static visualization skills. Mental rotation skills *would* likely be useful for comparing maps or molecules and their features to another map or molecule presented in a different orientation. These examples illustrate that performance benefits should be expected when spatial skills are applied in ways that fit particular problems; but when they do not fit, benefits are less likely to emerge.

Beyond these potential matches and mismatches, strategic choices can further complicate interactions between spatial skills and STEM performance. While tasks might present relationships for which particular spatial skills could be applied, people do not always rely on them (Hambrick et al. 2012; Schwartz and Black 1996). For example, Stieff (2007) employed a variant of the classic mental rotation task to assess differences between novice and expert performance on their recognition of chemistry molecules. Participants had to determine whether two molecules in a displayed pair were identical or mirror images. Novice chemistry students took longer to make their decisions when the second image was rotated at larger angles, analogous to findings from classic tasks involving the mental rotation of abstract (i.e., non-chemistry stimuli) blocks (Shepard and Metzler 1971). In contrast, for problems involving symmetrical molecules, experts did not show the traditionally obtained slowdowns. These data might suggest that expert chemists possess superior spatial skills allowing them to mentally manipulate objects so quickly that the amount of rotation between the pairs had little impact on their response times. However, complementary survey data supported a different explanation. While novices reported they were explicitly attempting to mentally rotate the images, chemistry experts reported

relying on an analytical heuristic derived from prior knowledge—comparing whether specific connected ions or molecules were identical. This strategy was generally fast and required no mental rotation to complete. Other results indicated that when experts were unable to use this non-spatial strategy, they relied upon mental rotation, obtaining typical response slowdowns with greater angular disparities. Thus, individuals can approach tasks in strategic ways that might engage spatial thinking to a greater or lesser degree (see also Stieff et al. 2012).

These findings indicate that spatial skills could be applied in effective or ineffective ways depending on an individual's understanding of, expectations about, and/or strategic approach toward completing a target task. Empirically evaluating situations for which performance proves less than optimal based on these influences serves as a crucial demonstration of potential challenges for successful and productive spatial thinking. To explore this issue, we studied a situation in the domain of chemistry in which spatial skills could be applied in a number of more or less useful ways. Participants were asked to view two simultaneous dynamic displays and subsequently explain the chemical processes underlying them (see Fig. 1 for a sample situation). One display presented a video depicting a physical laboratory experiment (e.g., of gas reaching equilibrium in two connected beakers), consistent with the kind of presentation that is readily observable in the real world. The other display presented a schematic animation depicting the movement of molecules in the same experimental situation, which is normally invisible to the naked eye. In general, novice chemistry students exhibit difficulty reasoning about particulate-level processes like those depicted in the animations (Abraham et al. 1994; Bodner 1991; Johnstone 1993; Novick and Nussbaum 1981). We were interested in whether students could integrate the information presented in the two displays to support an understanding of the underlying particulate-level chemistry principles conveyed by their pairing. This would involve understanding how the movement of molecules, as depicted in the animations, resulted in observable outcomes, as depicted in the videos. Most importantly, we were interested in how participants' spatial visualization skill might influence their use and comprehension of the displays.

Participants can take different approaches to tasks requiring the simultaneous consideration of multiple dynamic presentations. To document these types of approaches, we recorded participants' eye movements while they studied the two displays. Eye tracking provides an online measure of where participants look as they process material, with the assumption that their fixation locations are tied to attention and cognitive processing (Just and Carpenter 1976; Rayner 1998). Eye tracking has been used in this way to study participants' processing of complex spatial and multimedia presentations (e.g., Cohen and Hegarty 2007; Hegarty 1992; van Gog and Scheiter 2010), providing clues as to the visual features and content that support learners' thinking about and understanding of the materials. We used eye tracking to examine whether the influence of spatial abilities depends upon the observation strategies learners adopt as they attempt to understand and integrate the two displays.

Based on previous work and an informal task analysis, we derived two approaches that learners could adopt when presented with the multiple, complementary displays. One approach involves attempts to integrate their spatiotemporal features. This strategy would be reflected by eye movements transitioning back and forth between displays (see Johnson and Mayer 2012). In the current study, the movement of molecules in the animations informs a chemical explanation for the unfolding physical outcomes in the videos. Consider

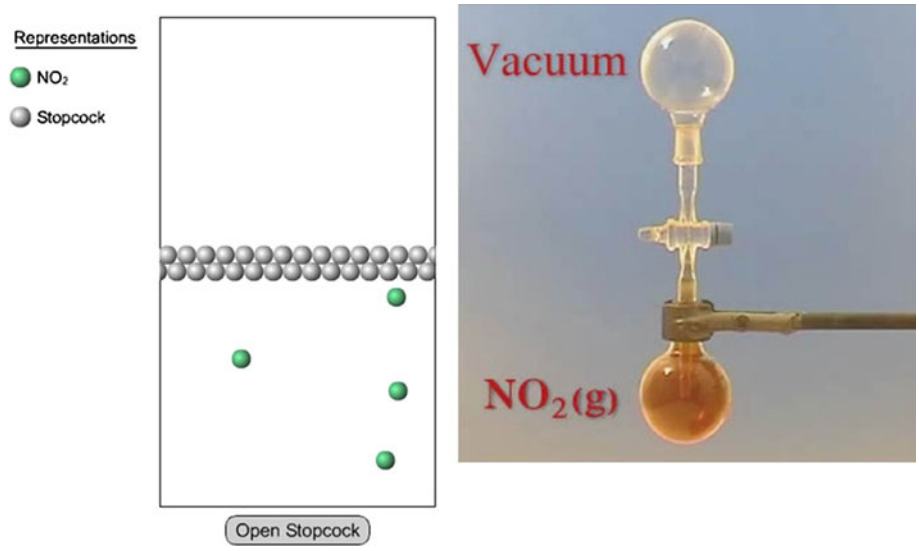

Fig. 1 Example screenshot of simultaneously presented particulate-level animation (*left*) and macroscopic-level video (*right*). More information on the development of these materials is available in Velázquez-Marcano et al. (2004)

the scenario in Fig. 1. The video demonstrates that when a stopcock is opened between two beakers, gas quickly fills both the gas-filled beaker and the empty beaker. The video does not, however, provide clear clues as to *why* this happens. Based purely on the video, participants may conclude that the vacuum actively pulls the gas up and that gas molecules speed up to fill an empty container, both of which are incorrect explanations. However, the matching animation displays the random motion of molecules at resting state which, when the barrier is opened, allows the molecules to continue spreading randomly throughout the greater volume. This random motion serves as the basis for a valid explanation, making the animation clearly important for comprehension (Williamson and Abraham 1995). But the animation *on its own* may also be insufficient to foster more complete understandings of the target concepts. The schematic presentation necessitates mapping the features of the animation (moving balls representing molecules, top and bottom halves of the screen) to the features of the video (NO_2 gas, the two beakers), and linking the type and timing of the enacted events. Thus, integration of the two displays is necessary to infer the causal principles underlying the lesson. This integration depends on the ability to construct a mental model of each situation and map the features between them (Mayer and Sims 1994), abilities that are likely tapped by spatial visualization skills. Thus, one prediction is that integration attempts should lead to better conceptual explanations if participants possess and employ the spatial abilities (i.e., spatial visualization) necessary to stitch together the two displays.

A second, less optimal approach to comprehending the displays involves predominantly relying on one display rather than integrating the two. Research on student use of complex displays has consistently demonstrated that learners strategically limit the amount of information they process based on what they consider relevant for completing a task (Hegarty 1992; Lowe 1999, 2004). Participants may thus focus more so on one display over the other, even though both are necessary for comprehension. How might this strategy interact with spatial visualization abilities? One consideration is that, in order to distinguish between information that is more or less important, participants must construct mental models from the spatiotemporal features of the displays (Hegarty et al. 2010; Lowe 2004). Individuals with higher spatial visualization abilities (HSVs) can more completely and efficiently construct these models from dynamic displays (Mayer and Sims 1994), which may facilitate their targeted deployment of attention to content they consider relevant to completing a task. Individuals with lower spatial visualization abilities (LSVs), in contrast, could exhibit less of a targeted search and thus would consider both displays in an effort to construct an understanding of the material. If participants engage in the types of focusing strategies outlined here, we hypothesize that LSVs should be less likely to restrict their focus to particular display content. In contrast, HSVs should be more willing or able to focus on the content they perceive to be task relevant. These claims lead to the counterintuitive hypothesis that HSVs will apply a sub-optimal strategy that limits the likelihood of integration when task instructions bias processing toward a single task-relevant display, while LSVs who do not restrict their search to this extent will rely on information from both displays. Data supporting these predictions would provide evidence for reduced performance despite greater spatial skills, as a function of task strategies and expectations (in contrast to what might be expected based on general processing propensities).

To assess the use of these strategies, and their relation to spatial visualization skills, we conducted two studies. In Study 1, before viewing the displays, participants with a range of spatial visualization skills were asked to predict the outcome of the experiment at the macroscopic level (i.e., what will happen in the videos). This type of prediction task, followed by viewing and explanation of an experiment, is representative of typical laboratory procedures involving prediction, observation, and explanation of scientific phenomena (Kearney 2004; Williamson and Jose 2009). This particular prediction task biases the importance of real-world considerations, as depicted in the videos, as task-relevant. We hypothesized that focusing to a greater degree on the videos would lead to sub-optimal performance, given that successful performance necessarily derives from integrating the videos with the animations. In Study 2, we manipulated whether participants received the prediction task, to determine whether removing it would mitigate any observed performance decrements. This directly tests whether the prediction task recruits processing strategies that, for some people, hinders comprehension of the target principles.

Study 1

In Study 1, participants viewed two complex dynamic displays, one meant to simulate a laboratory demonstration (at the macroscopic level) and one depicting the underlying movement of molecules (at the particulate level; Williamson and Abraham 1995; Williamson et al. 2012). Students are commonly asked to make predictions about chemical activity in classroom settings prior to looking at these types of instructional materials (Velázquez-Marcano et al. 2004), and the predictions are often made at the macroscopic level (i.e., what observation should one predict from an experimental procedure?). We adopted the same approach here, asking participants to make a prediction and only afterward presenting the two displays. In many cases, prediction tasks can be effectively leveraged

as a tool to elicit student understanding and foster deeper consideration of demonstrations (e.g., Kearney 2004). For the current project, because the prediction task focuses on macroscopic considerations, participants may be encouraged to sub-optimally focus on the macroscopic videos at the expense of the particulate animations. For example, a participant may predict that gas will move completely from the lower beaker to the upper beaker, perhaps believing that gas will invariably rise. When the simulation begins, the participant may try to observe whether their prediction was correct, spending time viewing the video showing the movement of gas in the beakers, rather than the animation displaying the molecular relationships underlying this event. Given that HSVs may be more efficient at focusing their attention than LSVs, we hypothesized that HSVs would actually be less likely than LSVs to consider both displays, focusing on the macroscopic presentation given the macroscopic task demands.

Method

Participants

Fifty general chemistry students at Texas A&M University, each in their first semester of college-level chemistry, completed the study in exchange for $15. To include students with a wide range of abilities, participants were recruited from high- and low-performing samples based on a course-administered chemistry quiz.[1] Five participants were excluded due to track loss and two participants' response data were incomplete, leaving 43 participants (30 females) with complete data.

Materials and apparatus

Eye tracker A Tobii T60 eye tracker was used to track eye movements while participants completed the comprehension tasks (see below). The eye tracker recorded the location and duration of eye fixations at a rate of 60 Hz using cameras embedded in a 17″ monitor with an estimated accuracy of 0.5° of visual angle.

Comprehension tasks Three "simulated experiment" trials were adapted from previous studies (Velázquez-Marcano et al. 2004) and used for the current study. Figure 1 displays an example trial screenshot. One trial focused on the diffusion of NO_2 gas into vacuum; a second focused on the diffusion of NO_2 gas into air; and the third trial focused on the response of liquid water to a vacuum. Each trial included a prediction screen that described the situation and asked participants to predict the end state by selecting one of five macroscopic-level images (see Fig. 2 for an example screen). The prediction task was critical, as we hypothesized it would encourage focus on the macroscopic level of representation.

Each trial included an explanation screen that asked participants to select an explanation from a set of multiple-choice (MC) items. For example, in the experiment involving NO_2 gas and air, participants were asked, "Why does the gas go up so slowly when there is air on top?" The most appropriate answer choice for this item was "there are frequent collisions between air and gas molecules," which derives from the particulate nature of matter (see Williamson and Abraham 1995), and is only apparent in the particulate-level animation. Less appropriate answers included, "there is no empty space [NO_2] needs to fill," "there is no force which pulls [NO_2] up," and "the heavier gas tends to stay at the bottom." Participants were also allowed to type in a response of their own. Each multiple-choice screen was followed by a justification screen. The justification screen provided a prompt for participants to "explain how you arrived at your answer. That is, provide a reason you chose your response, rather than the other options." This screen presented the answer choices at the top of the screen as a reminder, along with a text box for entering a typed justification. The justification screen afforded participants a second opportunity to provide a chemically appropriate answer. For example, given the prompt to explain why they did not choose the other answers for the previous simulation, a participant could state they believed the molecular explanation to be true (i.e., that molecules move randomly in all directions) but that the macroscopic explanation was more directly related to the question. Written justifications were coded for whether the explanation demonstrated acceptance of the molecular-level explanation. This justification opportunity was especially important given that some of the less appropriate answers were not completely incorrect. For example, the response choice from above, "there is no empty space [NO_2] needs to fill," is true to the extent that gases diffuse into empty space to fill their containers. However, this response does not explain *why* this diffusion process happens, nor *how* air molecules are involved, at a molecular level.[2] A participant could choose this less

[1] We found no effects of chemistry knowledge quiz scores on any of the outcome measures of interest. Controlling for scores on the chemistry quiz also did not account for any of the relationships discussed in the results section. Accordingly, we will not be discuss the quiz data further.

[2] There are other potential concerns with this answer, including the fact that the air does not take up *all* of the empty space in the container. Also, the answer appears to attribute intentionality to the NO_2 gas, which is inappropriate. While these issues are important, we were mainly interested in whether participants understood the target concept regarding molecular motion that is largely unrelated to these issues.

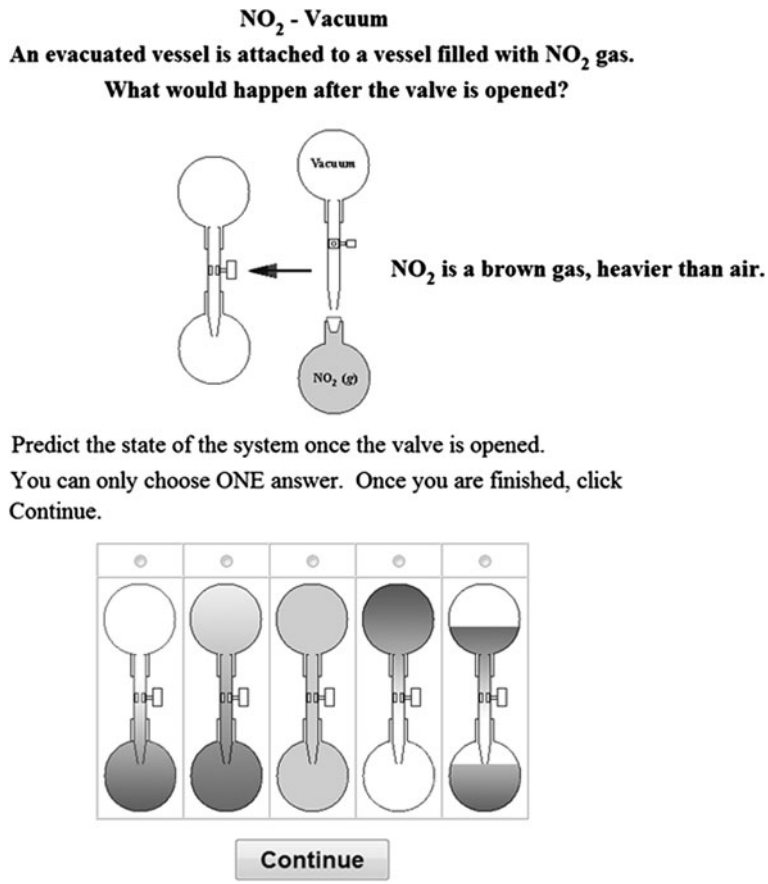

Fig. 2 Example screenshot of a prediction screen

appropriate but apparently correct answer even if they understood the underlying molecular-level chemistry. The written justification provided participants an opportunity to include chemically appropriate elaborations to supplement their multiple-choice selections.

Based on responses to the explanation and justification screens, we created two accuracy scores. The first score (Accuracy-MC) included only accuracy rates based on selection of the most appropriate answers from the MC options. The second score (Accuracy-Justified) included appropriate MC choices *as well as* less appropriate MC choices that were supplemented with chemically appropriate explanations in the written justifications.

Spatial ability tasks Participants completed four measures of spatial ability. Examples for three of the measures are shown in Fig. 3. The mental rotation test (MRT, Vandenburg and Kuse 1978) presents a three-dimensional block shape along with rotated block shapes that are either identical to or mirror images of the original shape. Participants select the image or images they believe are identical to the original shape. The Purdue visualization of rotations test (ROT, Bodner and Guay 1997) presents two three-dimensional objects that are identical but rotated. Participants also see a new object accompanied by a series of rotated target objects. Their task is to identify the target from the series that is rotated in relation to the new object in an analogous manner as the relations observed in the original pair of objects. Guay's visualization of viewpoints test (GVVT, Guay and McDaniels 1976) presents two identical, abstract three-dimensional objects presented from different perspectives. One object is surrounded by a "box" represented by dotted lines. Participants are asked to select the corner of the box of the first object from which the second object is being viewed. Each of these tasks involves the visualization and mental manipulation of an object in order to select an appropriate target. Thus, we reasoned that each of these tasks involves dynamic spatial visualization skills (Newcombe and Shipley in press). Participants also completed a hidden patterns test (Ekstrom et al. 1976), which is thought to measure intrinsic-static spatial skills. This test was included to ensure that the dynamic tasks of interest were differentiable from more static skills. All of the spatial tasks were presented electronically to allow for automatic scoring.

Eye tracking metrics To measure attention within and between the displays, we created two areas of interest

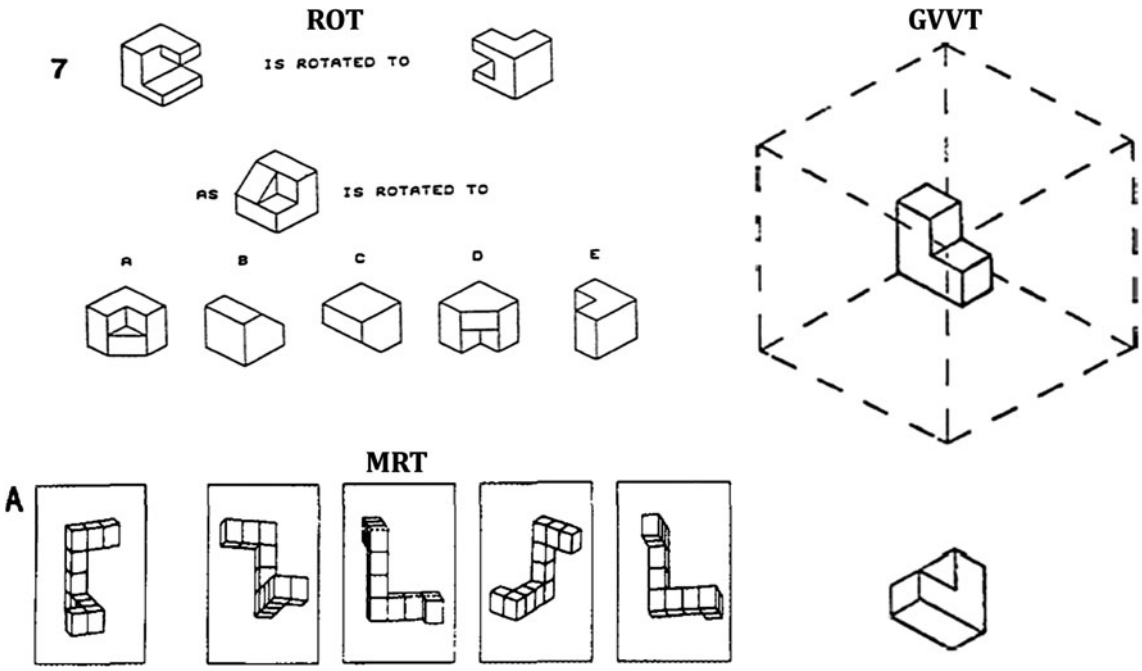

Fig. 3 Example items from dynamic spatial visualization tasks utilized in this experiment: ROT Purdue visualizations of rotations test. *MRT* mental rotations test, *GVVT* Guay's visualization of views test

(AOIs) surrounding the video and the animation for each trial. The AOIs around the videos were somewhat larger (29 % of the screen) than the AOIs around the animations (21 % of the screen), but these AOIs were consistent across trials and participants. We calculated the *total fixation duration* within each AOI. A fixation is identified when eye position is stable across a period of observations. Fixations are assumed to reflect participant attention to and processing of information near the eye position (Just and Carpenter 1976). In the current study, the total fixation duration within the animation AOIs provides an estimate of the amount of attention paid to the animations, similar to a measure of dwell time. We also created a relative measure by dividing total fixation duration on the animations by total fixation duration on both the animations and videos. This relative measure controls for time-on-task and for time spent outside of the AOIs to derive the relative attention paid toward the animations.

In addition to total fixation duration, we calculated the number of *transitions* between the two displays. A transition was coded whenever a fixation in one AOI was preceded by a fixation in the other AOI. For example, if a participant made five fixations on different parts of the video, then moved across the screen to make eight fixations on different parts of the animation, this would be coded as one transition. Movement to or from areas outside of the two AOIs did not count as transitions. Transitions were used to infer conceptual attempts at integrating the two displays (Johnson and Mayer 2012).

Procedure

All participants completed a block of spatial ability tasks and a block of comprehension tasks. In order to control for carryover effects, the order of these blocks was counterbalanced. All tasks were performed in the same session using the same computer.

Spatial ability tasks Participants were informed that they would be completing a series of "cognitive tasks" and were warned that some of the tasks may be difficult. The instructions for each task were presented via computer, and all instructions and tasks were self-paced. The presentation order of the four spatial tasks was randomized across participants.

Comprehension tasks Participants followed a standard procedure to calibrate the eye tracker. Next, they answered some basic demographic questions and then received instructions for the task. They were shown a brief example video as a demonstration of the type and timing of displays they would see, but were not asked to predict or explain the results. They were then informed that they would view several "simulated experiments" and would have to predict, observe, and explain the results. Each trial was then presented in counterbalanced order. For each trial, before viewing the displays, participants first completed the prediction screen. This required the selection of one of five multiple-choice macroscopic-level outcomes (see Fig. 2).

Then, participants viewed the simultaneous displays. The viewing period was partially self-paced. Participants saw a motionless screen as in Fig. 1. Using a mouse, they clicked a start button to begin viewing the simultaneous video and animation. The video and animation played for a set amount of time, with the participants unable to interrupt the presentation. At the end of each viewing, participants were allowed to click the button to view the presentation again, as many times as they wished, or they could press the space bar to move on. The system collected the number of button clicks for each participant as an estimate of the number of times they viewed the displays. When participants were finished viewing the displays, they were presented with the explanation screen, on which they selected a multiple-choice explanation and clicked a submit button with the mouse to continue. Then, they completed the justification screen by typing out a written justification for their response and clicking a submit button with the mouse to progress to the next trial. This procedure was repeated for the next two trials. After finishing the three trials, participants completed an unrelated chemistry task and, depending on counterbalancing, either moved on to the spatial ability tasks or were debriefed and thanked for their participation.

Results and discussion

Spatial ability score

We reasoned that the three dynamic measures of spatial ability represent a single spatial factor reflecting the ability to dynamically visualize and manipulate spatial information. Descriptive statistics for each of the spatial tasks are included in Table 1. As expected, there were positive correlations between the three measures (rs $> .38$, ps $< .05$). The measure of static spatial ability (the hidden patterns test) did not significantly correlate with any of the dynamic tasks (rs $< .17$) and was excluded from all subsequent analyses.[3] A factor analysis using a principle-axis factoring technique yielded a single factor which accounted for 66.63 % of the variance of the dynamic tasks. A standardized score was derived based on this factor analysis and used as a composite spatial visualization ability score. All subsequent analyses were conducted using this continuous measure of spatial visualization ability.

[3] If included in the factor analysis, scores on the hidden patterns test loaded on the same factor as the MRT, ROT, and GVVT. However, the factor loading was low (.20), and including hidden patterns scores reduced the fit of the model to account for only 51.53% of the variance. Given these results and the theoretical justification for considering hidden patterns as a separate construct, we excluded it from the spatial abilities factor score.

Spatial ability, attention to displays, and accuracy

Summary descriptive statistics for eye tracking and accuracy measures appear in Table 1.

With regard to total fixation duration, participants spent more time viewing the videos than the animations, $t(42) = 4.86$, $p < .001$. This finding may be due to a number of factors: The realistic features of the videos may have been more interesting or visually complex, or the introduction of a human element or cause in the videos (i.e., a human hand opening the stopcock) may have made them salient. Novice chemistry students may also be comfortable thinking at a macroscopic rather than particulate level of representation (Johnstone 1993). The video AOIs were also larger than the animation AOIs, so this comparison may be biased. We also noted a range of viewing times for both the animations and videos, which is not surprising given that participants were allowed to study the materials at their own pace. On average, participants viewed each trial 1.68 times (SD = .69). Thirteen participants viewed each trial only once. However, number of viewings and total viewing time did not relate to accuracy or spatial visualization ability, rs$(43) < .13$, suggesting that time-on-task factors are unlikely to play an obvious role in any relationships we report here. Nevertheless, as might be expected, participants who viewed the displays more times spent more total time viewing the videos, $r(43) = .58$, $p < .001$, and animations, $r(43) = .54$, $p < .001$, and made more transitions between the two, $r(43) = .56$, $p < .001$, than did participants who viewed the displays fewer times.

More importantly, we were interested in variance in the *relative* amount of attention to the two displays. We hypothesized that lower relative attention to the particulate animations (i.e., more exclusive focus on the macroscopic videos) would negatively correlate with accuracy rates. This hypothesis was supported for both measures of accuracy [MC: $r(43) = -.34$, $p = .03$; Justified: $r(43) = -.31$, $p = .04$], indicating that increased focus on the macroscopic videos was related to fewer chemically appropriate explanations. We also hypothesized that spatial visualization ability would be negatively correlated with relative attention to particulate animations if, as described earlier, HSVs were especially focused on the videos. This hypothesis was also supported, $r(43) = -.35$, $p = .02$, consistent with the previously proposed relationship between spatial ability and focus.

Finally, we hypothesized that spatial visualization ability would be negatively correlated with accuracy—in particular that HSVs would problematically attend to the macroscopic videos at the expense of attention to the particulate animations. This sub-optimal focus would be

Table 1 Descriptive statistics for both Study 1 and 2

	Prediction task instructions			No prediction task instructions		
	M	SD	Range	M	SD	Range
Study 1						
Accuracy-MC	.28	.27	.00:1.00	–	–	–
Accuracy-justified	.41	.33	.00:1.00	–	–	–
Total dur.—video	15.56	7.14	2.58:37.04	–	–	–
Total dur.—anim.	10.00	6.09	2.10:28.17	–	–	–
Relative anim. dur.	.41	.15	.08:.80	–	–	–
Transitions	13.24	6.97	4.33:37.33	–	–	–
MRT	16.98	7.78	4.00:38.00	–	–	–
GVVT	9.45	7.28	−1.67:24.00	–	–	–
ROT	56.28	21.19	15.00:100.00	–	–	–
Study 2						
Accuracy-MC	.24	.27	.00:1.00	.23	.25	.00:1.00
Accuracy-justified	.33	.27	.00:1.00	.35	.27	.00:1.00
Total dur.—video+	35.72	8.10	14.66:57.66	32.42	8.21	15.22:48.18
Total dur.—anim*	21.25	7.40	3.58:39.65	25.05	8.11	12.30:42.47
Relative anim. dur.*	.38	.12	.08:.72	.46	.13	.23:.73
Transitions^{+}	29.96	10.02	8.00:49.67	33.73	8.96	18.33:57.33
MRT	16.83	8.18	3.00:36.00	15.40	7.18	4.00:37.00
GVVT	10.77	6.80	−1.50:24.00	9.78	5.69	−3.83:20.50
ROT	59.39	20.16	25.00:95.00	55.75	20.24	20.00:85.00
Spatial vis. factor	.08	1.02	−1.78:2.14	−.08	.85	−2.04:1.45

Accuracy-MC accuracy rates only including multiple-choice responses, *Accuracy-Exp* accuracy rates including explanations, *MRT* mental rotation test, *GVVT* Guay's visualization of views test, *ROT* Purdue visualization of rotations test, *Spatial Vis Factor* composite score of all three spatial visualization tests, *—group difference significant at $p < .05$, $^{+}$—group difference marginally significant at $p < .10$

associated with performance decrements. This hypothesis was supported for both measures of accuracy [MC: $r(43) = -.43$, $p < .01$; Justified: $r(43) = -.37$, $p = .01$].

Transitions between displays

We expected that transitions between the two displays would reflect attempts at integration and would be positively related to performance. However, we found no evidence of correlations between transitions and either measure of accuracy, both rs$(43) = -.07$, nor a correlation between number of transitions and spatial visualization ability, $r(43) = .16$. We also tested whether the transitions might be more or less effective as a function of spatial ability. Hierarchical linear regression tested the independent contribution of a spatial ability X transitions interaction term after controlling for spatial ability and transitions. No evidence was obtained for this interaction for either Accuracy-MC ($\Delta r^2 = .02$) or Accuracy-Justified ($\Delta r^2 = .003$). Thus, transitions between displays exhibited no apparent relationship with accuracy.

Summary

Study 1 demonstrated that HSVs focused more exclusively on material that was related to prediction tasks than did LSVs. HSVs also scored lower on subsequent explanation tasks. This illustrates a situation for which spatial visualization ability is associated with sub-optimal processing and negatively related to performance. We believe that the prediction task encouraged attention to the macroscopic videos (or discouraged attention to the particulate animations), although it also could be the case that participants who preferred the videos attended to them for other reasons (e.g., perceived realism). In Study 2, a new set of participants viewed the displays and either engaged in the same prediction task or were not asked to make predictions. If the prediction task focused participants (particularly HSVs) on the macroscopic videos (or reduced attention to the particulate animations), then any asymmetric attention to the videos should be reduced when the prediction task is removed. We also tested whether removing the prediction task would encourage more transitions between displays and whether the resulting transitions would support conceptual understanding.

Study 2

Method

Participants

Eighty-nine general chemistry students from Texas A&M University completed the study in exchange for $15. To include students with a wide range of abilities, participants were recruited from high- and low-performing samples based on a course-administered chemistry quiz. These participants had also completed the Test of Logical Thinking (TOLT; Bunce and Hutchinson 1993), a measure of logical and conditional reasoning abilities as a classroom quiz.[4] Six participants' data were removed due to track loss and two participants had incomplete response data, leaving 81 participants with complete data (53 females).

Materials and procedure

The materials and procedure were identical to Study 1 with two exceptions. In Study 1, participants' overall viewing time for the displays was unconstrained (e.g., 13 participants only viewed each trial once, while other participants viewed individual trials up to four times). While these time-on-task factors did not relate to accuracy or spatial ability, we wanted to ensure that participants spent an adequate amount of time viewing each trial. To do this, we controlled the presentation of the displays so that each participant viewed each trial three times (which was greater than 1 SD more than the mean number of views from Study 1). Each comprehension task screen was presented in the same order as in Study 1, three times in a row before participants could move on to the explanation screen. Each participant completed the explanation and justification screens once for each set of displays.

For the second change in the procedure, participants were randomly assigned to either receive the prediction task ($n = 41$) or receive no prediction task ($n = 40$). Participants in the prediction condition received the same instructions as in Study 1. Participants in the no prediction condition were informed that they would be observing and explaining the results of simulated experiments, with no reference to the need to predict the results. These participants also did not receive the prediction screens for the task.

[4] In Study 2, we observed that spatial visualization ability was related to both the chemistry knowledge quiz, $t(79) = 6.48$, $p < .001$, and TOLT scores, $r(81) = .42$, $p < .001$. However, these variables did not relate significantly to the outcome variables of interest, and controlling for either of these variables did not account for the relationships between spatial visualization ability and the outcome variables. Accordingly, neither the chemistry knowledge quiz nor TOLT scores will be discussed further in Study 2.

Results

Spatial ability score

As in Study 1, there were positive correlations between the three dynamic spatial ability measures, $rs(81) > .35$, $ps < .01$. The hidden patterns test again did not relate significantly to the dynamic measures, $rs(81) < .17$, and was excluded from further analyses.[5] A factor analysis using a principle-axis factoring technique yielded a single factor that accounted for 65.97 % of the variance in the dynamic measures. Standardized scores were derived based on this factor analysis and used as a composite spatial visualization ability score.

Effect of prediction task

Descriptive statistics are provided in Table 1. We hypothesized that participants in the no prediction condition as compared to the prediction condition would spend more time on the particulate animations and less time on the macroscopic videos. This pattern was supported, as t tests confirmed a significant group difference in total fixation duration to animations, $t(79) = 2.20$, $p = .03$, and relative attention to animations, $t(79) = 2.75$, $p = .01$, with a marginal effect for total fixation duration to videos, $t(79) = -1.82$, $p = .07$. We also hypothesized that transitions would be more frequent in the no prediction as compared to the prediction condition. While this pattern held with regard to mean differences, the effect was marginally significant, $t(79) = 1.78$, $p = .08$. Despite these patterns of eye movement behavior, there was no effect of prediction task on either measure of accuracy ($ps > .10$).

Spatial ability, attention to displays, and accuracy

We next considered the role of spatial ability in each of the task conditions. In the prediction condition, there was a negative relationship between spatial visualization ability and relative attention to particulate animations, $r(41) = -.50$, $p < .01$, replicating the findings from Study 1. This relationship, however, was not observed with the no prediction group, $r(40) = .07$. A hierarchical linear regression confirmed a significant interaction between task condition and spatial ability, $\Delta r^2 = .07$, $F(1, 77) = 6.51$, $p = .01$, suggesting that spatial ability predicted greater attention to the videos, but only when preceded by predictions.

[5] As with Study 1, scores on the hidden patterns test loaded on the same factor as the MRT, ROT, and GVVT. However, the factor loading was low (.19), and including hidden patterns scores reduced the fit of the model to account for only 50.87 % of the variance. Thus, we again excluded it from the spatial abilities factor score.

Despite obtaining the expected pattern with regard to fixation duration, neither spatial ability nor relative attention to animations predicted accuracy in either condition (rs $< .18$). This pattern represents a failure to replicate the *negative* correlations found between spatial ability and accuracy measures following prediction tasks from Study 1.[6]

Transitions between displays

We first observed the total number of transitions as a factor of prediction condition and number of viewings. Participants made more transitions on the first view ($M = 39.39$, SD $= 13.03$), than the second ($M = 30.15$, SD $= 10.99$) or third view ($M = 26.45$, SD $= 11.57$), supported by a significant effect of viewing trial, $F(2, 156) = 48.10$, $p < .001$. There were also marginally more transitions in the no prediction than prediction condition (see Table 1), $F(1\ 78) = 2.74$, $p = .10$. Interestingly, on their first viewings, participants in the no prediction condition made more transitions ($M = 42.25$, SD $= 11.76$) than did participants in the prediction condition ($M = 36.53$, SD $= 13.74$), $t(78) = 2.00$, $p = .049$. This effect was not observed for second and third viewings (ts < 1.60).

Replicating Study 1, performance in the prediction condition revealed no relationship between transitions and either accuracy measure [MC: $r(41) = -.04$; Justified: $r(41) = -.03$]. In contrast, performance in the no prediction condition obtained significant positive correlations between transitions and accuracy [MC: $r(40) = .49$, $p < .01$; Justified: $r(40) = .41$, $p < .01$]. This interaction between task condition and transitions was confirmed for both accuracy measures [MC: $\Delta r^2 = .07$, $F(1, 77) = 6.35$, $p = .01$; Justified: $\Delta r^2 = .05$, $F(1, 77) = 6.51$, $p = .04$]. Integration attempts, as evidenced by transitions between displays, were related to accuracy, but only for participants who were not asked to generate predictions.

We also considered a possible relationship between spatial ability and transitions. Number of transitions did not correlate significantly with spatial visualization ability for either the prediction condition, $r(41) = .03$, or the no prediction condition, $r(40) = .18$. However, it may be that transitions were more strongly related to performance in one condition than the other. Recall that in the prediction condition, participants seemed to direct attention toward the video display, demonstrating a focusing strategy rather than an integration strategy. Given this pattern, the number of transitions may not reflect effective integration, even for HSVs. Consistent with this claim, as in Study 1, there was no interaction between spatial abilities and transitions for the accuracy measures ($\Delta r^2 < .08$, ps $> .09$). In the no prediction condition, given that transitions were marginally more common and *were* related to accuracy, they may reflect strategies aimed at integrating the displays. Effectively integrating the displays is likely dependent on spatial visualization skills (Mayer and Sims 1994) that reflect the ability to imagine and manipulate multiple mental models and relate the features of models to each other. If so, then transitions in the service of integration may be most effective for HSVs, given that integration attempts would be based on more complete and accurate mental models. In the no prediction condition, an interaction between spatial ability and transitions indeed obtained for the MC measure [MC: $\Delta r^2 = .12$, $F(1, 36) = 6.84$, $p = .01$; Justified: $\Delta r^2 = .01$, $F(1, 36) = .55$, $p = .47$]. These data provide some evidence that, while transitions were generally related to accuracy in the no prediction group, the transitions were most helpful for HSVs.

One important question is why this interactive effect held for MC accuracy, but was not significant for the accuracy measure when it included justifications. We note that the correlational patterns were similar for the two accuracy measures. To clearly illustrate this trend, we conducted a median split based on spatial ability. Looking specifically at the no prediction condition, LSVs revealed no significant relationship between transitions and accuracy [MC: $r(19) = .20$; Justified: $r(19) = 30$]. In contrast, HSVs showed a significant relationship between transitions and accuracy [MC: $r(21) = .72$, $p < .01$; Justified: $r(21) = .51$, $p = .02$]. We suspect that the relationships between transitions and spatial abilities were not categorically different for multiple-choice and justified measures, but were simply more statistically robust for the MC measure than the justified measure. A more speculative explanation for the difference in accuracy measures is that written justifications may reflect, in part, inferences generated *after* observing the displays, while initial MC responses are more responsive to online processing differences (i.e., transitions in the service of integrating the two displays; Rapp and Mensink 2011). When asked to justify a response, participants must reflect on why they chose their response *instead of* the other responses. Some participants may have observed the motion of the molecules in the animation, but not immediately infer that this movement accounted for results in the video, perhaps because of difficulties integrating the spatiotemporal relationships. Being asked to justify why the movement of molecules is not a correct explanation may offer

[6] It is surprising that we did not replicate this pattern. However, the replication of the eye tracking data indicates that the prediction task *did* direct attention toward the videos. Based on Study 1 and previous research (Williamson and Abraham 1995; Williamson et al. 2012), this processing strategy is not optimal. We also note that spatial visualization ability was not positively related to performance in the prediction condition, in contrast to prior research demonstrating generally positive relationships between spatial abilities and learning from visualizations.

participants another opportunity to make this inference. The current data do not speak directly to this speculation. However, the possibility that justifications may allow participants to overcome limitations associated with spatial visualizations skills is an intriguing issue for future study.

In general, the transition data are consistent with the notion that attempts to integrate the two displays were beneficial for understanding in the no prediction condition. In addition, the integration attempts were most beneficial for HSVs.

General discussion

The goal of the current project was to begin identifying ecologically representative conditions for which spatial skills might be used ineffectively, resulting in potential performance decrements. In Study 1, participants observed and provided scientific explanations for dynamic displays in a self-paced procedure. The displays depicted chemical phenomena occurring simultaneously at macroscopic (real-world video) and particulate (molecular animation) levels. Participants made a prediction at the macroscopic (real-world) level before observing the displays. This prediction served as a pre-study consideration intended to focus attention on the macroscopic level for the remainder of the task. In line with predictions, analyses of eye tracking and accuracy data showed that greater focus on the macroscopic videos, derived from the pre-study task, was associated with lower performance on measures of comprehension for the concepts in the displays. Spatial visualization ability was also related to focus on the macroscopic level and to subsequent performance. HSVs focused their attention more exclusively toward the macroscopic videos than toward both displays, which was problematic given that information from both displays is necessary for comprehension. LSVs, in contrast, did not show the same biased focus or associated performance decrements.

In Study 2, we tested whether these differential effects emerged specifically as a function of the focus instantiated by the pre-study prediction task. Participants viewed the same displays and completed the same comprehension tasks, with only half of the participants completing the pre-study prediction task. Eye tracking data verified that the prediction task guided attention: Participants who generated predictions focused more exclusively on the macroscopic video and made marginally fewer transitions between the two displays as compared to participants who were not asked to generate predictions. Spatial visualization abilities were again related to more exclusive focus on the macroscopic level, but only for participants who were asked to make predictions and not for participants who skipped that task. These results support the contention that participants' spatial visualization skills were strategically directed in response to the prediction task. Eliminating that task also made the benefits of an integration strategy more readily apparent, as participants who made more rather than fewer transitions between the two displays demonstrated better performance. This pattern was moderated by spatial visualization ability for one measure of accuracy, suggesting that the integration strategy was most effective for participants with the spatial visualization skills necessary to effectively imagine and manipulate information from the simultaneous displays.

The results of Studies 1 and 2 reveal that spatial visualization skills can be applied in effective and ineffective ways. HSVs focused on information promoted by the pre-study task, even though that focus limited the amount and type of information useful for understanding the concepts conveyed in the displays. One interpretation of this result is that participants believed they were focusing on relevant information based on the pre-task instructions, a strategy that is often useful for comprehending complex presentations (Hegarty et al. 2010; Lowe 2004). Spatial visualization abilities, then, can support focus on targeted information. For the current project, though, building a valid understanding of the displayed content required consideration of multiple displays, making a targeted focus less optimal. Study 2 demonstrated that HSVs more effectively considered multiple displays when the prediction pre-task was removed, supporting their performance on the comprehension task. Previous work has similarly shown HSV's success at integrating information from multiple modalities presented simultaneously (Mayer and Sims 1994). Spatial visualization skills can thus be applied in the service of at least two sets of processing strategies (e.g., differentiating relevant vs. irrelevant information, integrating models from multiple displays), and the application of these strategies can potentially support or limit comprehension.

We note that while the findings across both studies were largely consistent, in at least one case the findings diverged. In Study 1, participants' spatial ability and focus were related to accuracy, but the same pattern did not obtain in Study 2. It is not immediately clear why the accuracy patterns differed across the two studies even though the eye tracking patterns across them were similar, although there are several possibilities. In Study 1, participants viewed the displays at their own pace, but in Study 2, participants were required to view each display three times in succession, which resulted in longer viewing times on average. One possibility is that fixation durations and spatial abilities are most strongly associated with performance outcomes during initial viewings of displays (i.e., Study 1), but these factors decrease in importance, or other

factors become important, after repeated viewings (i.e., Study 2). (Recall that, on average, participants in Study 1 viewed the displays only one or two times). A related consideration is that total fixation duration and spatial abilities were more tied to performance when participants had control over how often they viewed the displays (Study 1), as compared to when they did not have control (Study 2). User control might influence the types of strategies and behaviors that participants enact when they are learning from simultaneous displays, as control has similarly been shown to influence student motivation toward engaging in tasks (Cordova and Lepper 1996). Each of the above possibilities suggests additional features of learning experiences (i.e., pacing, interactivity, perceived control) that could influence the application and success of spatial abilities in the service of comprehension.

Other than removal of the prediction task, we did not attempt to modify the displays to better serve as effective learning activities. But modifications could support student learning in a variety of relevant ways. For example, it may be possible to improve performance by adding attentional cues through narration or signaling (see Mayer 2001) or to present the different levels of representation sequentially, with some directive to integrate the two (see Williamson et al. 2012). These manipulations may help to direct participants toward integration, supporting attempts at understanding the relationships between molecular movement and the real-world displays of experimental simulations and visualizations.

The findings reported here have implications for STEM educators and researchers in considering the role of spatial skills when learning from presentations. Instructors who present simultaneous displays might expect students to engage in particular strategies (e.g., integrating the two displays) that will lead to intended outcomes, although based on task instructions, students may choose to employ other, potentially less effective or direct strategies (e.g., focusing on one display) to complete the tasks. Considering the particular features of any task and how it may encourage students to interact with instructional materials in different ways proves crucially important. In the current studies, pre-study prediction tasks, which are consistently used in STEM settings (Kearney 2004; Velázquez-Marcano et al. 2004), were associated with ineffective processing strategies and, in Study 1, with worse performance for the group possessing a relevant spatial skill (i.e., HSVs). Curricular developers should think carefully about how instructions and task demands might encourage unexpected approaches to learning. These considerations underlie the importance of identifying alignments and disconnects between task demands, and the allocation of spatial skills, as have been outlined in some taxonomies of spatial abilities (Newcombe and Shipley in press). We note that in accord with these considerations, care should be taken in generalizing the findings from our prediction manipulation to other prediction tasks or activities. Pre-study tasks often help individuals *more* effectively comprehend information (see Bransford and Schwartz 1999; Hegarty et al. 2003), particularly when the tasks elicit knowledge or strategies that align with the processes or activities necessary for successful performance. The purpose of the current project was to demonstrate cases in which visualization ability could be directed in ineffective ways when pre-study tasks encouraged attention to limited subsets of information. For the current studies, the focusing strategies encouraged by the prediction task did not align with the integrative processes most useful for understanding the chemical concepts. In contrast, if the information primed by pre-study tasks is particularly relevant or useful for a task, then this focus could certainly be beneficial.

Spatial skills are critical for thinking effectively about STEM domain content. A more complete understanding of the role of spatial skills in STEM learning, though, necessitates careful considerations of when, how, and why spatial skills may be employed for particular tasks (Uttal and Cohen 2012). The current project presents experimental data demonstrating that the effective application of spatial abilities can depend on the types of expectations and approaches that participants take in considering STEM visualizations. Future research should focus not only on whether individual differences in spatial skills are important for STEM learning, but how learners strategically employ spatial skills under different circumstances. Differential abilities, expectations, and tasks necessarily interact to influence learning outcomes.

Acknowledgments This work was funded by REESE grants #0907780 and #0908130 from the National Science Foundation to PIs Dr. David N. Rapp and Dr. Mary Jane Shultz.

References

Abraham MR, Williamson VM, Westbrook SL (1994) A cross-age study of the understanding of five chemistry concepts. J Res Sci Teach 31:147–165

Bodner G (1991) I have found you an argument: the conceptual knowledge of beginning chemistry graduate students. J Chem Educ 68:385–388. Retrieved from http://pubs.acs.org/doi/abs/10.1021/ed068p385

Bodner GM, Guay RB (1997) The Purdue visualization of rotations test. Chem Educ 2:1–18

Bransford JD, Schwartz DL (1999) Rethinking transfer: a simple proposal with multiple implications. Rev Res Educ 24:61–100

Brunyé TT, Taylor HA, Rapp DN, Spiro A (2006) Multimedia presentation of procedures: influences on memory for structure and content. Appl Cogn Psychol 20:917–940

Bunce DM, Hutchinson KD (1993) The use of the GALT (Group Assessment of Logical Thinking) as a prediction of academic success in college chemistry. J Chem Educ 70:183–187

Carroll JB (1993) Human cognitive abilities: a survey of factor analytic studies. Cambridge University Press Cambridge, New York

Cohen CA, Hegarty M (2007) Individual differences in use of external visualisations to perform an internal visualisation task. Appl Cogn Psychol 711:701–711

Cordova DI, Lepper MR (1996) Intrinsic motivation and the process of learning: beneficial effects of contextualization, personalization, and choice. J Educ Psychol 88:715–730

Ekstrom RB, French JW, Harmon HH (1976) Manual for kit of factor-referenced cognitive tests. Educational Testing Service, Princeton

Guay R, McDaniels RE (1976) The visualization of viewpoints. The Purdue Research Foundation, West Lafayette

Hambrick DZ, Libarkin JC, Petcovic HL, Baker KM, Elkins J, Callahan CN, Turner SP, Rench TA, LaDue ND (2012) A test of the circumvention of-limits hypothesis in scientific problem solving: the case of geological bedrock mapping. J Exp Psychol. doi:10.1037/a0025927

Hegarty M (1992) Mental animation: inferring motion from static diagrams of mechanical systems. J Exp Psychol Learn Mem Cogn 18:1084–1102

Hegarty M (2010) Components of spatial intelligence. Psychol Learn Motiv 52:265–297. doi:10.1016/S0079-7421(10)52007-3

Hegarty M, Waller D (2005) Individual differences in spatial abilities. In: Shah P, Miyake A (eds) The Cambridge handbook of visuospatial thinking. Cambridge University Press, New York, pp 121–167

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

Hegarty M, Canham MS, Fabrikant SI (2010) Thinking about the weather: how display salience and knowledge affect performance in a graphic inference task. J Exp Psychol Learn Mem Cogn 36:37–53

Höffler TN (2010) Spatial ability: its influence on learning with visualizations—a meta-analytic review. Educ Psychol Rev 22:245–269. doi:10.1007/s10648-010-9126-7

Johnson CI, Mayer RE (2012) An eye movement analysis of the spatial contiguity effect in multimedia learning. J Exp Psychol Appl. doi:10.1037/a0026923. (Advance online publication)

Johnstone A (1993) The development of chemistry teaching: a changing response to changing demand. J Chem Educ 70:701–705

Just M, Carpenter P (1976) Eye fixations and cognitive processes. Cogn Psychol 8:441–480. doi:10.1016/0010-0285(76)90015-3

Kali Y, Orion N (1996) Spatial abilities of high-school students in the perception of geologic structures. J Res Sci Teach 33:369–391

Kearney M (2004) Classroom use of multimedia-supported predict-observe-explain tasks in a social constructivist learning environment. Res Sci Educ 34:427–453. doi:10.1007/s11165-004-8795-y

Keehner M, Hegarty M, Cohen C, Khooshabeh P, Montello DR (2008) Spatial reasoning with external visualizations: what matters is what you see, not whether you interact. Cogn Sci 32:1099–1132. doi:10.1080/03640210801898177

Kozhevnikov M, Motes M, Hegarty M (2007) Spatial visualization in physics problem solving. Cogn Sci 31:549e579

Linn MC, Peterson AC (1985) Emergence and characterization of sex differences in spatial ability: a meta-analysis. Child Dev 56:1479–1498

Lowe RK (1999) Extracting information from an animation during complex visual learning. Eur J Psychol Educ 14:225–244

Lowe R (2004) Interrogation of dynamic visualization during learning. Learn Instr 14:257–274

Mayer RE (2001) Multimedia learning. Cambridge University Press, New York

Mayer RE, Sims VK (1994) For whom is a picture worth a thousand words? Extensions of a dual-coding theory of multimedia learning. J Educ Psychol 86:389–401. doi:10.1037/0022-0663.86.3.389

Newcombe NS, Shipley TF (2012) Thinking about spatial thinking: new typology, new assessments. In Gero JS (ed) Studying visual and spatial reasoning for design creativity. Springer, New York (in press)

Novick S, Nussbaum J (1981) Pupils' understanding of the particulate nature of matter: a cross-age study. Sci Educ 65:187–196

Pellegrino JW, Kail RV (1982) Process analyses of spatial aptitude. In: Sternberg RJ (ed) Advances in the psychology of human intelligence, vol 1. Erlbaum, Hillsdale, pp 311–365

Rapp DN, Mensink MC (2011) Focusing effects from online and offline reading tasks. In: McCrudden MT, Magliano JP, Schraw G (eds) Text relevance and learning from text. Information Age Publishing, Greenwich, pp 141–164

Rayner K (1998) Eye movements in reading and information processing: 20 years of research. Psychol Bull 124:372–422

Rochford K (1985) Spatial learning disabilities and underachievement among university anatomy students. Med Educ 19:13–26

Sanchez CA, Branaghan RJ (2008) The interaction of map resolution and spatial abilities on map learning. Int J Hum Comput Stud 67:475–481

Schwartz DL, Black Y (1996) Shuttling between depictive models and abstract rules: induction and fall-back. Cogn Sci 20:457–497

Shea D, Lubinski D, Benbow C (2001) Importance of assessing spatial ability in intellectually talented young adolescents: a 20-year longitudinal study. J Educ Psychol 93:604–614

Shepard RN, Metzler J (1971) Mental rotation of three-dimensional objects. Science 171:701–703

Stieff M (2007) Mental rotation and diagrammatic reasoning in science. Learn Instr 17:219–234. doi:10.1016/j.learninstruc.2007.01.012

Stieff M, Ryu M, Dixon BL, Hegarty M (in press) Problem solving strategies used by organic chemistry undergraduates. J Chem Educ

Thurstone LL (1938) Primary mental abilities. Psychom Monogr 1

Uttal DH, Cohen CA (2012) Spatial thinking and STEM education: when, why, and how? Psychol Learn Motiv 57:147–181. doi:10.1016/B978-0-12-394293-7.00004-2

van Gog T, Scheiter K (eds) (2010) Eye tracking as a tool to study and enhance multimedia learning [special issue]. Learn Instr 20

Vandenberg SG, Kuse AR (1978) Mental rotations, a group test of three-dimensional spatial visualization. Percept Mot Skills 47:599–601

Velázquez-Marcano AA, Williamson VM, Ashkenazi G, Tasker R, Williamson KC (2004) The use of video demonstrations and particulate animation in general chemistry. J Sci Educ Technol 13:315–323

Wai J, Lubinski D, Benbow CP (2009) Spatial ability for STEM domains: aligning over 50 years of cumulative psychological knowledge solidifies its importance. J Educ Psychol 101:817–835

Williamson VM, Abraham M (1995) The effects of computer animation on the particulate mental models of college chemistry students. J Res Sci Teach 32:521–534

Williamson VM, Jose TJ (2009) Using visualization techniques in chemistry teaching. In: Pienta NJ, Cooper MM, Greenbowe TJ (eds) Chemists guide to effective teaching, vol 2. Prentice Hall, Upper Saddle River

Williamson VM, Lane SM, Gilbreath T, Tasker R, Ashkenazi G, Williamson KC, Macfarlane RD (2012) The effect of viewing order of macroscopic and particulate visualizations on students' particulate explanations. J Chem Educ 89:979–987

Wu H, Shah P (2004) Exploring visuospatial thinking in chemistry learning. Sci Educ 88:465–492