Learning and Instruction 26 (2013) 12–21

Contents lists available at SciVerse ScienceDirect

Learning and Instruction

journal homepage: www.elsevier.com/locate/learninstruc

Beyond ball-and-stick: Students' processing of novel STEM visualizations

Scott R. Hinze [a,b,*], David N. Rapp [a,b], Vickie M. Williamson [c], Mary Jane Shultz [d], Ghislain Deslongchamps [e], Kenneth C. Williamson [f]

[a] School of Education & Social Policy, Northwestern University Evanston, 2120 Campus Drive, IL 60208, USA
[b] Department of Psychology, Northwestern University Evanston, 2120 Campus Drive, IL 60208, USA
[c] Department of Chemistry, Texas A&M University, College Station, TX 77843, USA
[d] Department of Chemistry, Tufts University, Medford, MA 02155, USA
[e] Department of Chemistry, University of New Brunswick, Fredericton, New Brunswick, E3B5A3, Canada
[f] Department of Construction Science, Texas A&M University, College Station, TX 77843, USA

ARTICLE INFO

Article history:
Received 31 August 2011
Received in revised form
8 December 2012
Accepted 11 December 2012

Keywords:
Visualization
Chemistry learning
Prior knowledge
Eye tracking

ABSTRACT

Students are frequently presented with novel visualizations introducing scientific concepts and processes normally unobservable to the naked eye. Despite being unfamiliar, students are expected to understand and employ the visualizations to solve problems. Domain experts exhibit more competency than novices when using complex visualizations, but less is known about how and when learners develop representational fluency. This project examined students' moment-by-moment adoption patterns for scientific visualizations. In a laboratory experiment, introductory-level organic chemistry students viewed familiar ball-and-stick and novel electrostatic potential map representations while solving chemistry problems. Eye movement patterns, verbal explanations, and individual difference analyses showed that students initially relied on familiar representations, particularly for difficult questions. However, as the task unfolded, students with more prior knowledge began relying upon the novel visualizations. These results indicate adoption and fluent use of visualizations is not given; rather, it is a function of prior knowledge and unfolding experience with presented content.

1. Introduction

1.1. Motivation for research

Scientific inquiry requires the consideration of concepts and objects that are often invisible to the naked eye. For example, researchers, students, and the lay public are frequently confronted with processes that occur at submicroscopic levels or at geologic time scales. A challenge for understanding these unobservable processes involves envisioning the components inherent to their events, and the effects that emerge from their interactions. Failure to comprehend these processes and effects can result in flawed inquiry activities, poor course grades, and scientific misconceptions. A variety of external representations are intended to help people understand unobservable scientific processes and effects (Gilbert, 2005; Gilbert, Reiner, & Nakhleh, 2008). In particular, *visualizations* (i.e. external visual representations) attempt to concretely anchor concepts using familiar features such as color, size, shape, and proximity to transform unobservable and abstract concepts into familiar and understandable symbols. Some of these visualizations have become ubiquitous in STEM fields (science, technology, engineering, and mathematics), including the ball-and-stick models found in chemistry classrooms, and choropleth maps provided in newspapers, journals, and textbooks (conveying, as examples, temperature gradients, neural activity, or electrostatic potential through color change).

These visualizations have been lauded as important tools for STEM learning, both for teaching scientific processes and concepts, and for helping students build *representational competence* or *fluency* (i.e., understandings of when and how to rely on different external representations, including the complementary or diverse inferences they may afford; Kozma & Russell, 1997). Representational competence is important because visualizations are not automatically converted into knowledge (see Rapp & Kurby, 2008) and learners may not know when or how to use a new visualization within a domain. For considering the implementation of visualizations in

* Corresponding author. Department of Psychology, Northwestern University Evanston, 2120 Campus Drive, IL 60208, USA. Tel.: +1 6169901076.
E-mail address: s-hinze@northwestern.edu (S.R. Hinze).

0959-4752/$ – see front matter © 2012 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.learninstruc.2012.12.002

textbooks, in classroom activities, and in popular press explanations of scientific phenomena, it is crucial to know (a) the nature of individuals' unfolding interactions with visualizations, (b) whether a student's ability influences the adoption of visualizations, and (c) the types of educational tasks that best support, and are best supported by, students' interactions with scientific visualizations. Each of these questions has implications for documenting and understanding students' developing representational competence.

1.2. Theoretical framework

Kozma and Russell (1997, 2005) described *representational competence* as a set of skills involving the comprehension and use of diverse representations. Representational competence requires, among other skills, the ability to differentiate the purposes of different representations and to understand when and why to use one representation over another. Far from being automatic, this competence necessitates learning how to analyze, generate, and explain the features of different types of external representations. Domain experts often possess these skills, gained through practice and experience, as evidenced by their ability to easily transition between representations based on the affordances of each representation for a particular situation. For example, expert chemists are able to consider what a molecule might look like from a variety of perspectives, across different nomenclatures and symbolic formats (Kozma, Chin, Russell, & Marx, 2000).

The skills underlying representational competence have mainly been documented through observations of expert performance and through explicit comparisons between expert and novice performance on comprehension tasks. Differences between these groups likely emerge because experts have more complete and coherent domain knowledge than novices, as well as more practice working with visualizations from their domains. In general, previous work indicates that instruction and practice with complex visualizations is necessary to help learners understand the underlying concepts conveyed in those visualizations (see Schwonke, Berthold, & Renkl, 2009). However, while these crucial supports have received substantial attention in classic and contemporary accounts of expert-novice performance (Chase & Simon, 1973; Chi, Feltovich, & Glaser, 1981; Haider & Frensch, 1999), there have been fewer empirical investigations of how competency emerges over time with regard to representation use. Identifying exactly when and how learners begin relying on novel visualizations offers a necessary analysis of how representational competence develops, which to date, has rarely been examined as learning unfolds (see Stieff, Hegarty, & Deslongchamps, 2011).

How are indicators of representational competence expressed when novices are first exposed to a novel visualization? Evaluating such expressions should help reveal when and how novices come to understand the affordances of novel visualizations relative to more familiar ones, and the types of decisions they make about when to apply visualizations to particular situations. Doing this necessitates identifying the developmental trajectories associated with representational fluency. With this approach as a goal, the current study examined performance on chemistry problem solving tasks to which participants could rely on either a novel visualization specifically useful for answering relevant questions, or a familiar visualization which, while informative, was less useful for answering the questions. This allowed for the examination of whether, when, and how learners would come to rely on unfamiliar but useful visualizations.

Previous research on representational competence and the comprehension of complex displays provides clues as to some of the factors relevant for novices' adoption of novel visualizations. Ainsworth (2006) noted two potential processes in such development: (1) individuals must understand the format of a representation, and (2) individuals must garner the relationship between the new representation and the domain of interest. The former involves identifying the particular features and operators necessary to decode a representation, while the latter involves generating inferences about the information conveyed in the representation and how the information fits into the larger content domain (see also Kozma & Russell, 2005).

Previous projects have considered the roles of these different processes, reporting that novices find it easier to interpret the format of a representation than to connect the underlying ideas conveyed in a representation to broader content understandings (e.g. Carpenter & Shah, 1998; Chi et al., 1981; Lowe, 1994). The two processes are also implicated by the finding that novices perform better on memory tasks that require retrieval of the surface features of representations than they do on tasks involving the integration and application of that information to solve problems (Mayer, 2001). These patterns of process-driven performance also have crucial implications for learning to adopt visualizations, given that reliance on novel visualizations depends on the tasks that individuals are asked to complete, and the types of understandings necessary to complete them. Because of their lack of familiarity, novices may attend to a novel visualization for tasks that only require basic understandings of format, such as identifying or comparing salient values or features in a display (Hegarty, Canham, & Fabrikant, 2010). In contrast, if a task requires constructing an inference that goes beyond encoding salient features to additionally extract and apply underlying principles to a variety of problems, novices may have difficulty adopting visualizations.

While task requirements can guide how and when novices rely on unfamiliar visualizations, the abilities and prior knowledge that individuals possess are also likely to moderate if and when they are willing and able to adopt novel visualizations. These individual differences can also help identify the particular strategies and activities that learners rely on to support their performance (Hegarty, 2010; Uttal & Cohen, 2012). Of particular interest for the current study, derived from previous findings on expertise and representational competence, was the role of prior knowledge. Prior knowledge can support learners' understandings of visualizations, especially when the materials are difficult to comprehend (e.g. Canham & Hegarty, 2010; Lowe, 1994). Consider that experts, who possess more prior knowledge in a domain than do novices, utilize visualizations successfully (see Cook, 2006; Gegenfurtner, Lehtinen, & Säljö, 2011 for reviews) by efficiently directing attention to relevant components of visualizations; novices, in contrast, show more indiscriminate patterns of attention (Haider & Frensch, 1999; Hegarty et al., 2010; Jarodzka, Scheiter, Gerjets, & van Gog, 2010; Lowe, 1994, 1999). Understanding how prior knowledge may facilitate selection decisions as students gain familiarity with visualizations would inform theoretical accounts of the development of representational fluency.

Chemical education presents a particularly relevant domain for these considerations, given the instructional challenge of presenting unobservable concepts and processes core to thinking about and reasoning in chemistry, and the emphasis that chemistry educators have placed on visualizations as a means of conveying relevant information. In chemistry classes, students exhibit substantial difficulty integrating macroscopic, symbolic, and submicroscopic levels of representation (Johnstone, 1993), and external visualizations are ostensibly useful for facilitating such understandings precisely by fostering representational competence (Williamson & Jose, 2009).

1.3. Overview of experiment and hypotheses

For the current project, we selected visualizations from the domain of chemistry, including a novel display that has enjoyed substantial growing support in chemistry textbooks and coursework, and another, more familiar display considered ubiquitous to the field. Electrostatic potential maps (EPMs) are emerging as a popular means of introducing students to core principles in chemistry, receiving endorsements from chemists and science educators (e.g., Sanger & Badger, 2001; Shusterman & Shusterman, 1997). EPMs present a color-coded representation of electron distribution in a relatively direct format (see Fig. 1). More familiar ball-and-stick models represent the shape of a molecule but in contrast to EPMs do not indicate the distribution of electrons. In the current experiment, we presented tasks for which EPMs should be useful given that information regarding electron distribution could be used to solve the problems. Information regarding individual atoms or bonds, as highlighted by ball-and-stick models, is less relevant for completing these tasks. The EPMs and ball-and-stick representations were presented simultaneously to determine on which display participants tended to rely. Presenting the representations simultaneously also aligns with textbook and classroom resource materials that, at times, offer multiple representations that students may consider, select, and/or integrate as they attempt to understand concepts and solve problems. Most importantly, this design allowed us to examine whether and when participants would adopt the EPMs over ball-and-stick models, as an indicator of developing familiarity with the novel EPMs and any emerging representational competence.

Participants' reliance on the visualizations was examined as they attempted to make basic identifications of and inferences about chemistry molecules. Identification questions required attention to the *format* of the visualizations, while inference questions required learners to understand chemical features of the molecules and make decisions about how these features would interact with other molecules or protons. Our first assumption was that problem-solving accuracy should be greater for identification than inference questions, because the more difficult inference questions require going beyond merely identifying features to additionally map connections between features and their chemical interactions. Our second assumption was that accuracy would be greater for participants with high as compared to low prior knowledge because prior knowledge of basic chemical concepts such as electronegativity and electron distribution should help support identifying and making inferences about chemistry features and processes. Inference questions required judgments about chemical interactions between positive and negative charges, which should be specifically supported by prior knowledge. In contrast, a basic understanding of the features of the visualizations may be enough to answer identification questions. Thus our first hypothesis, guided by the above assumptions, was that an interaction would moderate the above effects, with prior knowledge expected to exert larger effects on accuracy for inference than identification questions (Hypothesis 1).

Accuracy measures, though, provide limited insight into the processes that learners engage in as they attempt to solve problems and develop representational fluency (Rapp & van den Broek, 2005; Rapp, van den Broek, McMaster, Kendeou, & Espin, 2007). To assess developing competency with and adoption of the novel visualizations, we used an online eye tracking methodology to determine the degree of visual attention directed to the novel EPMs versus the standard ball-and-stick displays. We also observed whether any patterns of attention would change over the course of the experiment as participants practiced with the visualizations. We predicted a positive correlation between attention toward EPMs and accuracy, given that the EPMs were designed to be specifically useful for completing the task (Hypothesis 2). We also predicted that participants would rely more on the familiar ball-and-stick plots than the EPMs in many circumstances. These more familiar plots may support students' thinking about the particular atoms and molecules related to completing the task, despite the features displayed in them being less immediately useful than the relative charges conveyed by EPMs. Specifically, we hypothesized that attention toward EPMs would be greatest when the connections between EPMs and relevant concepts were best understood. Thus, we predicted that attention toward EPMs would be greater for identification than inference questions (Hypothesis 3a), and greater for participants with high than low prior knowledge (Hypothesis 3b). We also predicted prior knowledge would guide EPM adoption more strongly for inference than identification questions (Hypothesis 3c). These predictions align with the predictions earlier offered for performance accuracy.

Our eye tracking analysis also allowed us to examine whether developing familiarity with a novel visualization, over the course of several iterations of test questions, would increase reliance on that display. We hypothesized that participants would learn to selectively attend to EPMs over the course of practice (Hypothesis 4). Patterns consistent with this prediction would provide evidence of students acquiring representational fluency, which to date has rarely been demonstrated as learning unfolds (Stieff et al., 2011).

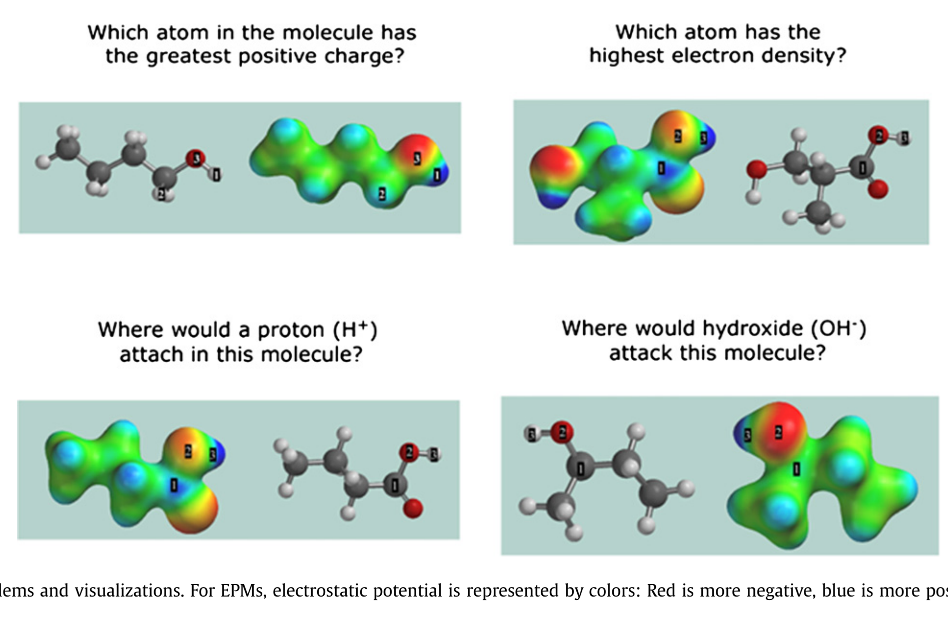

Fig. 1. Example problems and visualizations. For EPMs, electrostatic potential is represented by colors: Red is more negative, blue is more positive, green is neutral.

Additionally, Hypotheses 2–3 suggest a potential mediational model whereby prior knowledge is related to problem-solving accuracy by way of EPM use. Thus, we hypothesized that dwell time on EPMs during problem-solving may mediate the relationship between prior knowledge and accuracy (Hypothesis 5).

Finally, to complement the eye tracking measures, we collected verbal explanation data after participants solved the problems. Our analyses of these data focused on whether students explicitly reported using EPMs for problem-solving, to identify convergences between participants' explicit strategies and their eye movement data (see Van Gog, Paas, & Van Merrienboer, 2005). We predicted positive correlations between EPM use in explanations and visual attention to EPMs (Hypothesis 6). We also predicted that use of EPMs in verbal explanations would be more frequent for identification than inference questions (Hypothesis 7a), more frequent for high than low knowledge participants (Hypothesis 7b), and that prior knowledge would exhibit a stronger association with inference than with identification questions (Hypothesis 7c). These predicted effects align with the predictions offered earlier concerning eye movement and accuracy data.

Because high and low knowledge participants could differ in a variety of ways we also collected measures of other cognitive abilities that have been shown to covary with chemistry knowledge, and have been associated with learning. Chemistry knowledge and performance has been linked to reasoning abilities (Bunce & Hutchinson, 1993) and spatial ability (e.g. Wu & Shah, 2004), so we assessed these abilities as associated with performance on the current task. We also assessed differences in need for cognition (Cacioppo, Petty, & Kao, 1984) between high and low knowledge participants, to determine whether performance discrepancies might also be a function of learners' motivational tendencies. Each of these measures, while not core to the analysis, prove fruitful in determining student characteristics that may moderate the use of novel versus familiar visualizations.

2. Method

2.1. Participants

Students were recruited from a large organic chemistry course at Texas A&M University ($N = 225$). As confirmed by instructors at the institution, no formal instruction on the use of EPMs had been provided in class or in their general chemistry courses prior to the experiment. In contrast, instructors reported frequent use of ball-and-stick models. The models were used as graphical displays in lecture and as hands-on tools for use during laboratory exercises involving molecular geometry. Besides these implementations, students were encouraged, but not required, to purchase and use molecular model kits to support their independent study. High and low prior knowledge participants were recruited from this sample based on their scores on a pretest (see materials). A total of 35 students were selected from the course (19 high prior knowledge; 17 low prior knowledge), completed some portion of the experiment, and were compensated $20 for their time. Four participants were excluded for failing to complete both sessions of the study. One participant had unusable eye tracking data due to poor calibration and was also excluded, leaving a final sample of 30 participants (18 high prior knowledge; 12 low prior knowledge). The 30 participants consisted of 19 female and 11 male students with ages ranging between 19 and 21 years ($M = 19.47$, $SD = 0.68$).

2.2. Materials and apparatus

2.2.1. Pretest

As an in-class quiz, the 225 students answered 18 multiple choice questions regarding relevant chemistry concepts including the definition of electronegativity, the identification of oxidation states, and issues regarding molecular polarity. This pretest ($\alpha = 0.50$) served as the basis for recruiting high knowledge (top 1/3 of scores) and low knowledge (bottom 1/3 of scores) participants. Scores on the pretest ranged from 3 to 16 ($M = 9.23$, $SD = 2.53$, median = 9), with a relatively normal distribution (Skewness = 0.05, SE = 0.17). Based on this range, participants correctly answering 7 questions or less ($n = 57$; pretest $M = 5.92$, $SD = 1.19$) were recruited as low prior knowledge (referred to as LPK) and participants correctly answering 11 questions or more ($n = 57$; pretest $M = 12.32$, $SD = 1.43$) were recruited as high prior knowledge (referred to as HPK). From this recruitment sample, 35 students, as described previously, participated in the experiment proper.

2.2.2. Eye tracker

For portions of the experiment, a Tobii T60 eye tracker was used to track eye movements while participants completed the provided problem set. 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 degrees of visual angle at distances ranging between 50 and 80 cm.

2.2.3. Instructional materials

Participants were introduced to EPMs with a 391-word multimedia text presented via computer. The text explained key concepts such as the attraction of opposite charges and the role of electronegativity differences in facilitating this process. These concepts were illustrated by the use of two examples: the ionic bond between Na^+ and Cl^- based on their widely different electronegativity, and the less extreme case of the bonds in H_2O. EPMs for each of these examples were presented to show positive (blue), negative (red) and neutral (green) charges. Students did not practice using EPMs during the presentation of the instructional materials, and read the materials at their own pace.

2.2.4. Problem set

The main task in this experiment asked participants to use ball-and-stick and/or EPM representations to answer questions about molecules. Examples of the questions and representations are provided in Fig. 1. All problems consisted of a question displayed at the top of the screen along with simultaneous presentation of both EPM and ball-and-stick representations of a molecule. Half of the items presented the EPM on the left and half presented the EPM on the right. There were a total of six different molecules depicting various chemical structures. Three answer choices (with only one being correct) numbered 1–3 were presented on atoms within each representation. Four questions were created for each of the six molecules, for a total of 24 items. Two of the questions for each molecule required participants to identify an atom that met certain basic criteria (i.e. greatest positive charge, highest electron density); we refer to these as *Identification* questions (see top panels of Fig. 1). The two other questions for each molecule required participants to determine where a positive or negative charge would be attracted in the molecule; we refer to these as *Inference* questions (see bottom panels of Fig. 1).

2.2.5. Cognitive ability measures

Participants completed a battery of tasks presented via computer including tests of their reasoning and spatial abilities, and need for cognition. The purpose of this analysis was to determine whether prior knowledge differences were related to other cognitive or motivational factors. The reasoning abilities task was an electronic version of the Test of Logical Thinking (TOLT, Tobin & Capie, 1981), a 10-question test containing a variety of reasoning problems, which has been positively correlated with successful chemistry performance (Bunce & Hutchinson, 1993). Spatial abilities were assessed using measures intended to tap a variety of spatial components. Two tests measured the speed of mental rotation including the Card Rotation Test (CRT; Ekstrom, French, Harman, & Dermen, 1976) and the Mental Rotation Test (MRT; Vandenberg & Kuse, 1978). Complex visualization abilities were measured with the Purdue Visualization of Rotations Test (ROT; Bodner & Guay, 1997) and Guay's Visualization of Viewpoints (GVVT; Guay & McDaniels, 1976). Gestalt identification ability was assessed using the Hidden Patterns Test (HPT; Ekstrom et al., 1976). Participants also completed the 18 question Need for Cognition scale (NFC; Cacioppo et al., 1984) as a measure of motivation.

2.3. Eye tracking metrics

To examine participants' eye movement patterns during their completion of the problem sets, we concentrated on the location and duration of *fixations*, derived from an automated analysis using an I-VT filter (for details on fixation filters, see Komogortsev, Gobert, Jayarathna, Koh, & Gowda, 2010). Individual fixations occur when the eye stops for a period of time (in our analysis, a minimum of 60 ms) in a relatively stable spatial location (in our analysis, within 0.5 degrees of visual angle). Two videos are provided online (one LPK participant and one HPK participant) which demonstrate representative patterns of eye-movements. The *x* and *y* coordinates of each fixation (seen in the videos as individual numbered circles) were recorded along with the duration of each fixation (represented in the videos as the size of each circle). The video examples can be viewed at http://www.sesp.northwestern.edu/learning-sciences/eyetracking-sample.html.

While data were collected on fixations for the fully presented screen, we were mainly interested in fixations to the areas containing either the EPM or the ball-and-stick representation for each item. The size of these two Areas of Interest (AOIs) was held constant across all trials and the two representations were similar in size ($\sim$20% of the area of the screen for each AOI). We calculated measures related to the number and length of *fixations* within an AOI and the total amount of dwell time spent within an AOI for a given trial. While some researchers have proposed that fixation duration (i.e. the average length of fixations) and number of fixations can represent different constructs (Schwonke et al., 2009), these measures were highly correlated in our data suggesting convergence. Across item types and AOIs, total dwell time (i.e. the total fixation duration in an AOI) was related to measures of the average fixation duration (r's = 0.41–0.65) and the number of fixations within an AOI (r's = 0.83–0.97). These correlations suggest that multiple metrics reflecting the amount of time viewing an AOI would likely be redundant. As these measures were obtained while participants were attempting to answer questions, we considered the total dwell time metric to reflect the relative amount of attention paid toward each visualization in service of answering the question (see Just & Carpenter, 1976). Thus, longer dwell time on EPMs (relative to dwell time on ball-and-stick models) was interpreted as more reliance on EPMs for answering a question.

2.4. Explanation coding

While eye movement data quantify the amount of time participants spend viewing a representation during problem-solving, viewing time does not necessarily reflect the *use* of that representation to answer a question. Rather, longer fixation time on a representation may reflect difficulty in comprehension, or even have little relationship to a participant's explicit strategies. Verbal explanations can serve as additional evidence for the types of information that participants explicitly utilized to answer questions (Kendeou & van den Broek, 2005; Van Gog et al., 2005). After their response and eye-tracking data were collected for each problem, participants were asked to explain their response with the prompt, "please explain how you arrived at your answer out loud." These responses were transcribed and coded based on the type of information used as support for the explanation. In order to determine which representations participants relied upon, all responses were coded based on content referencing the two visual representations. Explicit strategies could reference features or components of the EPMs, the ball-and-stick models, or both, reflecting the types of visual and conceptual information that participants used to answer the questions. One set of codes identified references related to EPMs, including colors of the EPMs (red, blue, orange, etc.) or mentions of electronegativity or electron density (i.e. the concepts targeted by the EPMs). We inferred that explanations referencing the concepts or features of the EPMs reflected reliance on these representations for answering the questions. The other set of codes identified references related to ball-and-stick models, including individual atoms within the molecule (i.e. any response that specifically named an atom or molecular subgroup within a molecule), or references to atomic bonds or valence electrons. These chemical features are most directly communicated by the ball-and-stick plots, since the EPMs present space-filling models that obscure individual atoms or bonds. We inferred that explanations referencing these concepts and features reflected reliance on ball-and-stick representations for answering the questions. Each set of codes could potentially be applied to participants' explanations, including instances in which participants referenced both representations. To obtain a measure of reliability, two of the authors, one a chemistry specialist, coded all responses from six participants (20%). The Cohen's Kappa for our results was 0.85 indicating strong agreement. The remaining responses were coded by the first author. To specifically examine any developing representational fluency with the unfamiliar EPMs, we focused our analyses on the proportion of explanations referencing *only* the novel EPM displays.

2.5. Procedure

The pretest and TOLT were completed prior to participation in the experiment as part of the participants' organic chemistry coursework. All other activities took place in two counterbalanced laboratory sessions (one eye tracking session and one individual differences session), with each participant's second session occurring a week after the first. Informed consent was provided at the beginning of the first session, and payment was provided upon completion of the second session.

In the eye tracking session, after calibrating the eye tracker, participants read through the instructional materials on EPMs at their own pace. Participants then freely viewed two representations of all six molecules without question prompts. This viewing session lasted 5 s per molecule, and as a baseline presentation served to make the participants generally aware of the molecules they would be working with and the format of the representations. It also served as a check for whether participants' attention would be generally drawn toward one representation over the other (e.g. participants could simply be more intrigued by the colorful EPM plots). Next, participants answered the 24 items in the problem set at their own pace, presented in random order. Each item was presented for an unlimited duration, and participants were asked to provide an answer to each item both with a keypress (1, 2, or 3 on the number pad) and verbally (as recorded by an experimenter and with a microphone), and then to press the spacebar to advance to the next screen. Participants then rated their confidence and provided an oral explanation of why they chose each answer. The explanation was collected immediately after a participant's response was recorded, to avoid potential interference between verbal and visual-spatial processing when queried in a simultaneous fashion (Ericsson & Simon, 1980). Participants were prompted to "please explain how you arrived at your answer out loud." The explanation period was self-paced. If a participant did not verbalize an explanation, they were reminded by the experimenter using the same prompt statement. After answering all questions in this fashion, the eye tracking session was completed.

The individual differences session was completed in small groups in a private computer lab. Participants were seated at a computer that displayed each of the cognitive abilities tasks in random order. Participants completed these tasks sequentially, using a mouse to select responses and advance screens at their own pace.

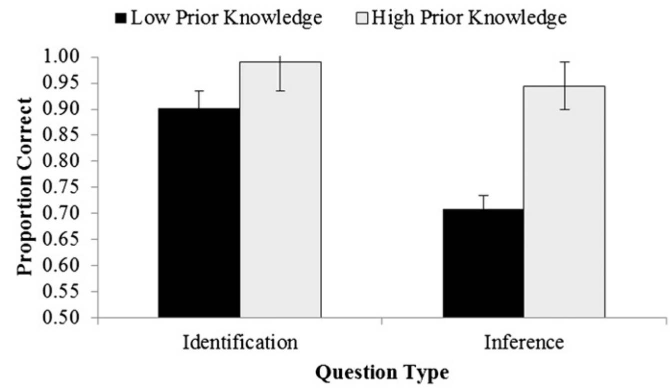

Fig. 2. Mean response accuracy based on prior knowledge and question type. Error bars in all figures represent ±1 SEM.

3. Results

3.1. Prior knowledge and other individual differences

Because students were selected based on high and low prior knowledge, it was important to determine whether students in these two groups might also be distinguished as a function of other cognitive abilities or demographic variables. The two prior knowledge groups did not differ based on measures of spatial ability, age, gender composition, or Need for Cognition (p's > 0.10). On the TOLT, a measure of logical and conditional reasoning abilities (Tobin & Capie, 1981), HPKs scored higher ($M = 7.89$ out of 10, SD $= 1.53$) than LPKs ($M = 6.25$, SD $= 2.70$) [$t(28) = 2.12$, $p = 0.04$, $d = 0.75$]. Thus, the effects of prior knowledge may relate to differences in reasoning abilities, but not with the other factors of interest tested. Below we present data based only on the prior knowledge groups, since this was the ability variable of greatest interest. We note here that we also performed correlational analyses relating spatial abilities and need for cognition with performance and eye-tracking, but given limited and inconsistent results (in fact, only the hidden patterns test was significantly related to accuracy for inference questions), we do not discuss these factors further.

3.2. Baseline preferences

Total dwell time was calculated based on data collected when students were asked to view all of the molecules for 5 s without question prompts. These data, calculated for ball-and-stick and EPM AOIs, reflect whether students were, *a priori*, differentially drawn to one type of representation. Participants spent an equivalent amount of time perusing ball-and-stick ($M = 2.37$ s, SD $= 0.35$) and EPM representations ($M = 2.30$ s, SD $= 0.39$), with no main effect or interaction with prior knowledge (F's < 1). Thus any observed bias toward a particular representation during problem-solving could not be attributed to an inherent preference, but rather should emerge from attempts to answer the test questions.

3.3. Accuracy

Response accuracy data are shown in Fig. 2. We expected that accuracy would be lowest for inference questions, lower for LPKs than HPKs, and that an interaction would emerge for which prior knowledge would be most important for inference questions (Hypothesis 1). These assumptions, and the hypothesis, were supported. Inference questions were more difficult ($M = 0.85$, SD $= 0.22$) than identification questions ($M = 0.96$, SD $= 0.12$) overall, $F(1, 28) = 10.71$, $p = 0.003$, $\eta_p^2 = 0.28$. LPKs also performed worse ($M = 0.81$, SD $= 0.19$) than HPKs ($M = 0.97$, SD $= 0.05$) overall, $F(1, 28) = 12.11$, $p = 0.002$, $\eta_p^2 = 0.30$. And these main effects were qualified by the predicted interaction, $F(1, 28) = 4.06$, $p = 0.05$, $\eta_p^2 = 0.13$: A larger accuracy difference was observed between the prior knowledge groups for inference, $t(28) = 3.25$, $p = 0.003$, $d = 1.11$, as compared to identification questions, $t(28) = 2.09$, $p = 0.05$, $d = 0.69$. Thus, the prior knowledge variable was most important for accuracy on inference questions.

3.4. Eye tracking of visualization use

We next tested the predictions regarding EPM adoption. We predicted that EPM adoption would be related to higher accuracy (Hypothesis 2). We also predicted that EPMs would be more fully adopted for identification than inference questions (Hypothesis 3a), and that HPKs would more fully adopt EPMs than would LPKs (Hypothesis 3b). Finally, we predicted an interaction whereby the effect of prior knowledge would be strongest for inference questions as compared to identification questions (Hypothesis 3c). To address these predictions, we calculated a single measure of relative EPM use by dividing the total dwell time on the EPM representation by the total dwell time on either the ball-and-stick or the EPM representation. Thus, scores above 0.50 indicate greater reliance and scores below 0.50 indicate less reliance on EPM than ball-and-stick representations. This measure helps to control for time on task, reflecting simply whether there was a bias toward one type of visualization over the other.

First, we confirmed that EPM adoption was useful (Hypothesis 2): Greater relative viewing time to EPM representations was significantly associated with higher accuracy ($r = 0.60$, $p < 0.001$). This relationship was significant for inference questions ($r = 0.57$, $p = 0.001$), but not for identification questions ($r = 0.20$, $p > 0.10$).

As predicted, representation use varied as a function of prior knowledge and question type (Hypotheses 3a and 3b; see Fig. 3). LPKs relied less on EPMs ($M = 0.45$, SD $= 0.12$) than did HPKs ($M = 0.53$, SD $= 0.10$), $F(1, 28) = 4.03$, $p = 0.05$, $n_p^2 = 0.13$, and both HPKs and LPKs relied less on EPMs for the inference questions ($M = 0.46$, SD $= 0.14$) than for the identification questions ($M = 0.55$, SD $= 0.11$), $F(1, 28) = 24.43$, $p < 0.001$, $n_p^2 = 0.47$. However, the interaction between prior knowledge and question type did not reach significance (Hypothesis 3c), $F(1, 28) = 2.60$, $p = 0.12$, $n_p^2 = 0.09$.

While HPKs were more willing and able to use EPMs to answer both identification and inference questions, aggregating across trials obscures whether practice might have influenced participants' developing reliance over the course of the experiment. Since

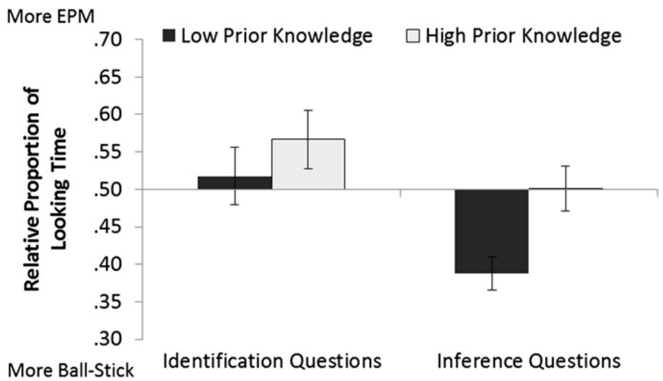

Fig. 3. Mean relative use of visualizations based on prior knowledge and question type. Higher proportion scores represent higher relative dwell time toward EPMs.

the EPMs were novel, students might only come to rely on them rather than the more familiar ball-and-stick depictions after extended exposure and practice. Thus, we next tested whether EPM adoption would increase with practice (Hypothesis 4), by assessing relative representation use across the six practice trials.

For identification questions, as seen in the left panel of Fig. 4, there was no main effect of Practice or Prior Knowledge (Fs < 1). However, a significant quadratic trend was obtained for the interaction between Practice and Prior Knowledge, $F(1, 28) = 4.47$, $p = 0.04$, $n_p^2 = 0.14$. LPKs did not significantly change their use of visualizations with practice, $F(1, 11) < 1$, while HPKs showed an initial increase in adoption of EPMs followed by a leveling off of reliance as practice proceeded, $F(1, 17) = 4.71$, $p = 0.05$, $n_p^2 = 0.22$. For inference questions, there was a main effect of Practice, $F(5, 140) = 4.39$, $p = 0.001$, $n_p^2 = 0.14$, and a main effect of Prior Knowledge, $F(1, 28) = 4.69$, $p = 0.04$, $n_p^2 = 0.14$. As shown in the right panel of Fig. 4, the effect of Practice differed for HPKs and LPKs as confirmed by a linear interaction, $F(1, 28) = 9.12$, $p = 0.01$, $n_p^2 = 0.25$. LPKs continued to view ball-and-stick representations more than EPMs throughout the experiment, $F(1, 11) < 1$; in contrast, HPKs showed a linear increase in EPM adoption, $F(1,17) = 23.35$, $p < 0.001$, $n_p^2 = 0.58$. In sum, HPKs and LPKs both initially relied on the familiar ball-and-stick representation. However, only HPKs learned to rely more on the novel EPMs over the course of the experiment.

We also conducted analyses using absolute, rather than relative, dwell times such that representation type (EPM, ball-and-stick) varied along with prior knowledge group (LPK, HPK) and question type (identification, inference). Overall, the patterns overlapped with the relative data, revealing a significant 3-way interaction, $F(1, 28) = 5.99$, $p = 0.02$, $\eta_p^2 = 0.18$. In addition, there was a marginal main effect of prior knowledge on total time on task, suggesting that LPKs spent somewhat longer answering questions overall, $F(1, 28) = 3.09$, $p = 0.09$, $\eta_p^2 = 0.10$. Thus, the relative measures discussed previously serve not only to simplify the interpretation of this 3-way interaction by combining the two representation types, but also control for marginal differences in overall response time.

3.5. Mediation analysis

These results explicate relationships between eye movement patterns, prior knowledge, and performance, suggesting that eye movement patterns may be responsible for the advantage exhibited by HPKs over LPKs on problem-solving accuracy. This mediational model (Hypothesis 5) was tested using stepwise multiple regressions, collapsing across question-type. In the model, binary Prior Knowledge was used as the initial predictor variable, relative EPM use was used as the mediator, and accuracy was the outcome variable. First, linear regression confirmed that Prior Knowledge predicted Accuracy, $\beta = 0.55$, $t = 3.48$, $p = 0.002$. Second, it was confirmed that Prior Knowledge predicted the mediator of relative EPM use, $\beta = 0.36$, $t = 2.01$, $p = 0.05$. Third, both EPM use and Prior Knowledge were entered to predict accuracy. This test confirmed that EPM use independently predicted accuracy, $\beta = 0.47$, $t = 3.17$, $p = 0.004$. After including EPM use in this last model, the relationship between Prior Knowledge and Accuracy was reduced, but still significant, $\beta = 0.39$, $t = 2.62$, $p = 0.01$. Since this relationship was reduced, we used a Sobel Test to assess the significance of the mediation, which was significant based on a one-tailed test, $t = 1.81$, SE $= 0.03$, $p = 0.045$. These results suggest that eye movement patterns partially mediated the relationship between Prior Knowledge and Accuracy.

3.6. Verbal explanations and use of representations

Finally, we tested the prediction that the patterns observed with verbal explanations would converge with the eye-tracking measures. (One participant's explanation data were unusable due to an audio recording error.) If, as argued, longer dwell times to EPMs reflected their adoption and use for problem-solving, then longer dwell times to EPMs should be associated with more explicit references to EPMs for explanations (Hypothesis 6). Consistent with this hypothesis, higher proportions of EPM-related explanations were associated with longer relative dwell times to EPMs for both identification ($r = 0.52$, $p = 0.01$) and inference questions ($r = 0.43$, $p = 0.02$). EPM-related explanations were not significantly associated with accuracy on identification questions ($r = 0.12$), but were associated with greater accuracy on inference questions ($r = 0.46$, $p = 0.01$), demonstrating similar relationships as the eye tracking data (see Hypothesis 2 above). These data suggest that participants' references to the features of EPMs (e.g. color) or to the conceptual information communicated by these displays (e.g. electron density) were related to their attention to EPMs and also to more accurate

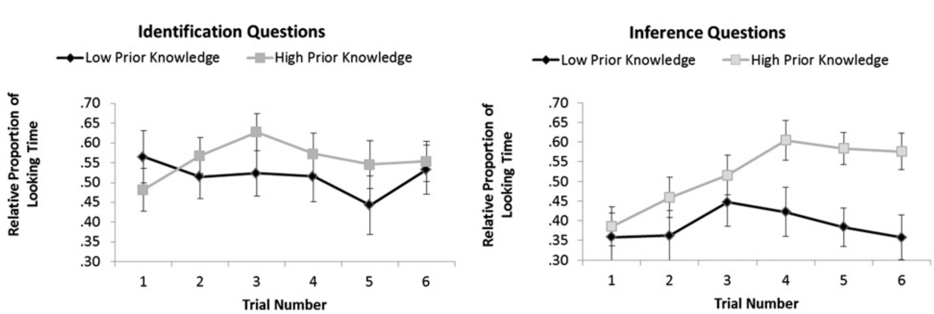

Fig. 4. Mean relative use of visualizations over practice. Higher proportion scores represent higher relative dwell time toward EPMs.

responses. The data converge on the conclusion that EPM adoption was successful for problem-solving, especially with regard to the difficult inference questions.

We next examined the role of question-type and prior knowledge in EPM-related explanations. We predicted that EPM-related explanations would obtain effects similar to eye-tracking metrics (i.e. main effects of prior knowledge and question-type and a possible interaction between the two variables, Hypotheses 7a–c). A mixed ANOVA demonstrated the predicted main effect of question-type (Hypothesis 7a), with a higher rate of EPM-related explanations for identification questions ($M = 0.52$, SD = 0.34) than for inference questions ($M = 0.37$, SD = 0.33), $F(1, 27) = 9.69$, $p = 0.004$, $\eta_p^2 = 0.26$. However, prior knowledge did not obtain a main effect (Hypothesis 7b), $F(1, 27) = 1.38$, $p = 0.25$, even though HPKs provided numerically more EPM-related explanations (identification: $M = 0.52$, SD = 0.35; inference: $M = 0.44$, SD = 0.33) than did LPKs (identification: $M = 0.46$, SD = 0.37; inference: $M = 0.23$, SD = 0.30). The interaction between prior knowledge and question-type was not significant (Hypothesis 7c), $F(1, 27) = 2.12$, $p = 0.16$.

In sum, analyses of the content of explanations largely demonstrated convergence between eye gaze data and participants' explicit justifications for their responses. Explanations involving components of EPMs were associated with longer viewing of EPMs and higher accuracy for inference questions. EPM-related explanations were more common for identification questions, for which the EPMs are most obviously relevant. The relationships between prior knowledge and explanations were similar in direction to the effects of prior knowledge on accuracy and eye gaze data, but were not statistically significant.

4. Discussion

This study explored how novice students come to adopt and rely on novel visualizations for solving problems, directly motivated by questions concerning the nature of representational fluency. The problems we tested were designed to be best solved using unfamiliar models, and less easily solved using more familiar models. Multiple methods were employed to assess whether adoption of the unfamiliar models would be successful, and to evaluate which factors might influence whether and how successfully participants readily adopted their use. Generally, the data indicated that such adoption *was* useful for problem-solving success. Eye tracking and explanation measures demonstrated that attention to and explicit use of EPMs was related to higher accuracy (Hypotheses 2 and 6).

More centrally, we were interested in factors that influence learners' developing reliance on novel representations. Based on broader claims about the comprehension and use of external representations, we predicted and identified roles for both question-type and prior knowledge. Participants were more likely to adopt the novel visualizations for answering questions that required identifying the format and features of a display than they were for answering questions that required generating inferences (Hypotheses 3a and 7a). This finding aligns with models and accounts that contend that drawing inferences from an external representation proves difficult, and thus is less likely to spontaneously occur, than identifying the surface-based features of an external representation (Ainsworth, 2006; Carpenter & Shah, 1998; Kozma & Russell, 2005). In the current project, participants were able to understand how the EPMs could be useful for answering identification questions about basic features, given that such identifications are relatively easy with practice. However, because drawing inferences from EPMs proves more difficult and less practiced, many learners may avoid using them to complete inferential tasks. This explanation suggests that students' adoption or avoidance of EPMs is driven, in part, by the difficulty of comprehending unfamiliar visualizations, even though their designs are specifically intended, and often lauded by designers, as supporting learning.

We also found that prior knowledge facilitated successful adoption of EPMs, as reflected by accuracy (Hypothesis 1) and eye tracking data (Hypothesis 3b), but not explanations (Hypothesis 7b). This finding is consistent with previous results showing that experts tend to more efficiently process external representations than do novices (Gegenfurtner et al., 2011; Haider & Frensch, 1999; Lowe, 1994). Prior knowledge also interacted with question-type for accuracy outcomes (Hypothesis 1), consistent with the view that knowledge proves especially helpful for tasks requiring inferences about content, but less crucial for interpretations of display features (Ainsworth, 2006; Mayer, 2001). This interaction indicates that, even when the features of a representation are clear, domain knowledge plays a necessary supportive role for understanding the relevance of these features to the domain of interest. Specifically with regard to representational competence, Kozma and Russell (2005) proposed a set of developmental stages that aligns with this obtained interaction. According to their account, novices move from more surface level understandings, termed "syntactic use," to deeper considerations of underlying meanings and constructs, termed "semantic use." The developmental skills necessary to successfully answer interpretation and inference questions from the current study fall under these "syntactic use" and "semantic use" categories, respectively. Prior knowledge, along with extended experience, fostered students' performance on inferential "semantic use" questions, while participants were better prepared to complete "syntactic use" activities for answering identification questions. These findings then, in line with the accounts and data culled from the extant literature, elucidate features that prove crucial for the development and application of representational competence.

Determining what these features specifically are is important, because while it is clear that experts and novices differ in their representational fluency (see Kozma et al., 2000), much less is known about how novices make representational decisions within a domain as they learn. The data presented here suggest that any developing representational competence relies not only on practice with visualizations, but also on task and learner characteristics during early learning and comprehension experiences. Participants in the current study did not immediately adopt novel EPM representations when asked to make inferences about chemical interactions, relying instead on more familiar ball-and-stick representations. Interestingly, participants with greater prior knowledge began to effectively apply the EPMs on inference problems with practice; low prior knowledge participants, in contrast, maintained their reliance on ball-and-stick representations and were more likely to answer questions incorrectly. Domain knowledge seems to facilitate early adoption and use of novel representations. Thus, not only can fluent representation use reflect differences in expertise (Kozma & Russell, 1997, 2005), but basic chemistry knowledge, even among relative novices, may facilitate the development of representational fluency.

Two caveats are worth noting regarding the above observations and interpretations. First, our measure of chemistry knowledge demonstrated relatively low internal consistency. This may be a consequence of an attempt to obtain a broad survey of general chemistry knowledge, such that pretest questions tapped different sub-types of chemistry knowledge or skills. However, this measurement concern was likely alleviated by the selection of extreme groups from a large sample, making it less likely that high- and low-knowledge participants were falsely assigned into the wrong group. Second, performance on the identification questions demonstrated high levels of accuracy rates for nearly all participants. This may have resulted in ceiling effects, with the effects of certain variables being weakened by a reduction in the potential variance in performance. Specifically, prior knowledge and use of EPMs (as measured by dwell time or explanations) may have demonstrated larger relationships with identification accuracy if performance was not so near ceiling levels. Subsequent investigations might benefit from implementing more variety in the types of questions, and the difficulty level of those questions, in assessments of participant performance.

As with most experiences, people rely on what they already know. The current experiment was no exception, as participants often defaulted to utilizing familiar ball-and-stick models to answer questions about chemistry processes. Successful learning, though, necessitates the consideration of new sources of information, to test alternative and unfamiliar perspectives and concepts. Students best able to adopt novel visualizations and answer questions related to them possessed knowledge about chemistry and general reasoning skills that might have (a) motivated a willingness or interest in considering new visualizations, (b) freed up necessary cognitive resources to contemplate unfamiliar representations, (c) fostered adaptive strategies for handling the test questions, or perhaps some combination of these processes. Thus, representational fluency was encouraged by the novel visualizations, but mainly for students who had existing knowledge, skills, or strategies that fostered taking representational chances. Our study does not evaluate the validity of these different mechanisms. Future work might attempt to distinguish between these possibilities by motivating students to focus on particular visualizations (e.g., through extrinsic rewards), further increasing or decreasing task difficulty, including dual task activities designed to overload cognitive resources, or training students on methods of answering test questions. These manipulations could determine whether the patterns emerging for participants with high prior knowledge might emerge more generally across a variety of students or conditions. The current project, though, offers a useful launching point for considering the crucial role of prior knowledge and reasoning skills in building representational competency.

While we have identified factors that might influence adoption of novel representations, this study has not identified or tested manipulations that may help remediate or support such adoption. One suggestion from the current data is that learners may benefit from opportunities to put novel visualizations to use. However, given the critical role of prior knowledge, only some learners may benefit from this practice. One intriguing possibility is that when students experience difficulty adopting relevant representations, that difficulty may offer insight into their gaps in knowledge, and provide an opportunity to revisit basic concepts assessed by target questions. In other words, if students are unable to select a representation as useful, even after practice, this may reflect that the student does not understand the underlying chemistry concepts that the representation relies upon for successful use. Researchers or instructors may then remediate these knowledge gaps, using the novel visualizations as an instructional tool or subsequent formative measure of understanding. Future research would be necessary to determine if these selection and adoption strategies could be used to diagnose and remediate knowledge gaps.

One interesting outcome of the experiment was that while prior knowledge, and to a lesser extent reasoning ability, were related to performance and representation use, spatial abilities measures did not exhibit similarly strong relationships. It may seem especially surprising that spatial abilities did not influence performance given the role of spatial abilities in the comprehension of other complex visualizations and multimedia (e.g., Cohen & Hegarty, 2007; Hannus & Hyona, 1999; Huk, 2006; Kalyuga, Ayres, Chandler, & Sweller, 2003; Mayer & Sims, 1994). It is important, however, to consider the particular task demands with which the students were engaged here. Participants were asked to reason about relative charges within a single static molecule, not requiring rotation to visualize their components. According to the typology offered by Newcombe and Shipley (in press), this task could be considered "static" in that participants were required to disembed components within space but were not required to think dynamically about those components or perform mental transformations with them (see also Schnotz, 2001). Given the mapping of the task to this typology, it seems reasonable that dynamic measures such as mental rotation or visualization of viewpoints would not strongly predict performance. These results should not be taken to suggest that spatial abilities cannot influence the development and application of representational competence more generally. Even for EPM use, we could imagine tasks that would necessarily be solved involving more dynamic imagery. For instance, students might be asked to predict the interaction of two EPMs in which the molecules could be rotated to connect areas of high and low electronegativity. Regardless of the potential role for spatial ability to moderate other tasks, the current data support the idea that spatial abilities need not mediate all types of visual tasks or visualization use, but rather that spatial skills may be more or less influential depending on specific task demands or learner approaches (Hegarty, 2010; Hinze, Williamson, Shultz, Williamson, & Rapp, in press; Stieff, 2007).

Visualizations, along with other supports offered by educational technologies (Bruce & Levin, 1997), can have great appeal for instructional designers given their affordances for conveying complex or otherwise abstract information. However, students appear to require practice with novel visualizations, and may approach these visualizations in disparate ways as a function of individual differences that are informed by both formal and informal experiences. The degree to which researchers, students, and laypeople understand when and how to utilize external representations, or whether they infer anything beyond the contents of those activities to generate novel hypotheses, theorems, and inventions, is not obvious. Logical justifications or arguments concerning the development of visualizations, and the advertised popularity of particular representations in textbooks, newspapers, and classroom materials, should not replace empirical evaluations of whether and how individuals come to rely on and benefit from experiences with them. Only by understanding when, how, and why different learners are able to adopt complex visualizations, and how competency with the displays unfolds over time, will we be able to fully realize their pedagogical potential.

Acknowledgments

This research was funded by REESE grant #0908130 from the National Science Foundation, David N. Rapp and Mary Jane Shultz Co-Principal Investigators.

References

Ainsworth, S. (2006). DeFT: a conceptual framework for considering learning with multiple representations. *Learning and Instruction, 16*, 183–198. http://dx.doi.org/10.1016/j.learninstruc.2006.03.001.

Bodner, G. M., & Guay, R. B. (1997). The purdue visualization of rotations test. *The Chemical Educator, 2*, 1–18. http://dx.doi.org/10.1333/s00897970138a.

Bruce, B. C., & Levin, J. A. (1997). Educational technology: media for inquiry, communication, construction, and expression. *Journal of Educational Computing Research, 17*, 79–102. http://dx.doi.org/10.2190/7HPQ-4F3X-8M8Y-TVCA.

Bunce, D. M., & Hutchinson, K. D. (1993). The use of the GALT (group assessment of logical thinking) as a prediction of academic success in college chemistry. *Journal of Chemical Education, 70*, 183–187. http://dx.doi.org/10.1021/ed070p183.

Cacioppo, J. T., Petty, R. E., & Kao, C. F. (1984). The efficient assessment of need for cognition. *Journal of Personality Assessment, 48*, 306–307. http://dx.doi.org/10.1207/s15327752jpa4803_13.

Canham, M., & Hegarty, M. (2010). Effects of knowledge and display design on comprehension of complex graphics. *Learning and Instruction, 20*, 155–166. http://dx.doi.org/10.1016/j.learninstruc.2009.02.014.

Carpenter, P. A., & Shah, P. (1998). A model of the perceptual and conceptual processes in graph comprehension. *Journal of Experimental Psychology: Applied, 4*, 75–100. http://dx.doi.org/10.1037/1076-898X.4.2.75.

Chase, W. G., & Simon, H. A. (1973). Perception in chess. *Cognitive Psychology, 4*, 55–82. http://dx.doi.org/10.1016/0010-0285(73)90004-2.

Chi, M. T. H., Feltovich, P., & Glaser, R. (1981). Categorization and representation of physics problems by experts and novices. *Cognitive Science, 5*, 121–152. http://dx.doi.org/10.1207/s15516709cog0502_2.

Cohen, C. A., & Hegarty, M. (2007). Individual differences in use of external visualizations to perform an internal visualization task. *Applied Cognitive Psychology, 21*, 701–711. http://dx.doi.org/10.1002/acp.1344.

Cook, M. P. (2006). Visual representations in science education: the influence of prior knowledge and cognitive load theory on instructional design principles. *Learning, 60*, 1073–1091. http://dx.doi.org/10.1002/sce.20164.

Ekstrom, R. B., French, J. W., Harman, H. H., & Dermen, D. (1976). *Manual for kit of factor-referenced cognitive tests. Educational testing service*. New Jersey: Princeton.

Ericsson, K., & Simon, H. A. (1980). Verbal reports as data. *Psychological Review, 87*, 215–251. http://dx.doi.org/10.1037/0033-295X.87.3.215.

Gegenfurtner, A., Lehtinen, E., & Säljö,, R. (2011). Expertise differences in the comprehension of visualizations: a meta-analysis of eye-tracking research in professional domains. *Educational Psychology Review, 23*, 523–552. http://dx.doi.org/10.1007/s10648-011-9174-7.

Gilbert, J. K. (2005). *Visualization in science education*. Dordrecht, The Netherlands: Springer.

Gilbert, J. K., Reiner, M., & Nakhleh, M. (2008). *Visualization: Theory and practice in science education*. United Kingdom: Springer.

Guay, R., & McDaniels, R. E. (1976). *The visualization of viewpoints*. West Lafayette, IN: The Purdue Research Foundation.

Haider, H., & Frensch, P. A. (1999). Eye movement during skill acquisition: more evidence for the information reduction hypothesis. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25*, 172–190. http://dx.doi.org/10.1037/0278-7393.25.1.172.

Hannus, M., & Hyona, J. (1999). Utilization of illustrations during learning of science textbook passages among low and high ability children. *Contemporary Educational Psychology, 24*, 95–123. http://dx.doi.org/10.1006/ceps.1998.0987.

Hegarty, M. (2010). Components of spatial intelligence. In B. H. Ross (Ed.), *The psychology of learning and motivation, Vol. 52* (pp. 265–297). San Diego: Academic Press.

Hegarty, M., Canham, M. S., & Fabrikant, S. I. (2010). Thinking about the weather: how display salience and knowledge affect performance in a graphic inference task. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 36*, 37–53. http://dx.doi.org/10.1037/a0017683.

Hinze, S. R., Willamson, V. M., Shultz, M. J., Willamson, K. C., Deslongchamps, G., & Rapp, D. N. When do spatial abilities support student comprehension of STEM visualizations? *Cognitive Processes*, in press.

Huk, T. (2006). Who benefits from learning with 3D models? The case of spatial ability. *Journal of Computer Assisted Learning, 22*, 392–404. http://dx.doi.org/10.1111/j.1365-2729.2006.00180.x.

Jarodzka, H., Scheiter, K., Gerjets, P., & van Gog, T. (2010). In the eyes of the beholder: how experts and novices interpret dynamic stimuli. *Learning and Instruction, 20*, 146–154. http://dx.doi.org/10.1016/j.learninstruc.2009.02.019.

Johnstone, A. H. (1993). The development of chemistry teaching: a changing response to changing demand. *Journal of Chemical Education, 70*, 701–705. http://dx.doi.org/10.1021/ed070p701.

Just, M. A., & Carpenter, P. A. (1976). Eye fixations and cognitive processes. *Cognitive Psychology, 8*, 441–480. http://dx.doi.org/10.1016/0010-0285(76)90015-3.

Kalyuga, S., Ayres, P., Chandler, P., & Sweller, J. (2003). The expertise reversal effect. *Educational Psychologist, 38*, 23–31. http://dx.doi.org/10.1207/S15326985EP3801_4.

Kendeou, P., & van den Broek. (2005). The effects of readers' misconceptions on comprehension of scientific text. *Journal of Educational Psychology, 97*, 235–245. http://dx.doi.org/10.1037/0022-0663.97.2.235.

Komogortsev, O. V., Gobert, D. V., Jayarathna, S., Koh, D., & Gowda, S. (2010). Standardization of automated analyses of oculomotor fixation and saccadic behaviors. *IEEE Transactions on Biomedical Engineering, 57*, 2635–2645.

Kozma, R. B., Chin, E., Russell, J., & Marx, N. (2000). The role of representations and tools in the chemistry laboratory and their implications for chemistry learning. *Journal of the Learning Sciences, 9*, 105–143. http://dx.doi.org/10.1207/s15327809jls0902_1.

Kozma, R. B., & Russell, J. (1997). Multimedia and understanding: expert and novice responses to different representations of chemical phenomena. *Journal of Research in Science Teaching, 34*, 949–968. http://dx.doi.org/10.1002/(SICI)1098-2736(199711)34:9<949, AID-TEA7>3.0.CO;2-U.

Kozma, R. B., & Russell, J. (2005). Students becoming chemists: developing representational competence. In J. K. Gilbert (Ed.), *Visualization in science education* (pp. 121–146). United Kingdom: Springer.

Lowe, R. K. (1994). Selectivity in diagrams: reading beyond the lines. *Educational Psychology, 14*, 467–491. http://dx.doi.org/10.1080/0144341940140408.

Lowe, R. K. (1999). Extracting information from animation during complex visual learning. *European Journal of Psychology of Education, 14*, 225–244. http://dx.doi.org/10.1007/BF03172967.

Mayer, R. E. (2001). *Multimedia learning*. New York: Cambridge University Press.

Mayer, R. E., & Sims, V. K. (1994). For whom is a picture worth a thousand words? Extensions of a dual-coding theory of multimedia learning. *Journal of Educational Psychology, 86*, 389–401. http://dx.doi.org/10.1037/0022-0663.86.3.389.

Newcombe, N. S., & Shipley, T. F. Thinking about spatial thinking: new typology, new assessments. In J. S. Gero (ed.), *Studying visual and spatial reasoning for design creativity*. United Kingdom: Springer, in press.

Rapp, D. N., & Kurby, C. A. (2008). The 'ins' and 'outs' of learning: internal representations and external visualizations. In J. K. Gilbert, M. Reiner, & M. Nakhleh (Eds.), *Visualization: Theory and practice in science education* (pp. 29–52). United Kingdom: Springer.

Rapp, D. N., & van den Broek, P. (2005). Dynamic text comprehension: an integrative view of reading. *Current Directions in Psychological Science, 14*, 276–279. http://dx.doi.org/10.1111/j.0963-7214.2005.00380.x.

Rapp, D. N., van den Broek, P., McMaster, K. L., Kendeou, P., & Espin, C. A. (2007). Higher-order comprehension processes in struggling readers: a perspective for research and intervention. *Scientific Studies of Reading, 11*, 289–312. http://dx.doi.org/10.1080/10888430701530417.

Sanger, M. J., & Badger, S. M., II (2001). Using computer-based visualization strategies to improve students' understanding of molecular polarity and miscibility. *Journal of Chemical Education, 78*, 1412–1416. http://dx.doi.org/10.1021/ed078p1412.

Schnotz, W. (2001). Sign systems, technologies and the acquisition of knowledge. In J.-F. Rouet, J. J. Levonen, & A. Biardeau (Eds.), *Multimedia learning: Cognitive and instructional issues* (pp. 9–30). Amsterdam: Pergamon.

Schwonke, R., Berthold, K., & Renkl, A. (2009). How external representations are used and how they can be made more useful. *Applied Cognitive Psychology, 23*, 1227–1243. http://dx.doi.org/10.1002/acp.1526.

Shusterman, G. P., & Shusterman, A. J. (1997). Teaching chemistry with electron density models. *Journal of Chemical Education, 74*, 771–776. http://dx.doi.org/10.1021/ed074p771.

Stieff, M. (2007). Mental rotation and diagrammatic reasoning in science. *Learning and Instruction, 17*, 219–234. http://dx.doi.org/10.1016/j.learninstruc.2007.01.012.

Stieff, M., Hegarty, M., & Deslongchamps, G. (2011). Identifying representational competence with multi-representational displays. *Cognition & Instruction, 29*, 123–145. http://dx.doi.org/10.1080/07370008.2010.507318.

Tobin, K., & Capie, W. (1981). Development and validation of a group test of logical thinking. *Educational and Psychological Measurement, 41*, 413–424. http://dx.doi.org/10.1177/001316448104100220.

Uttal, D. H., & Cohen, C. A. (2012). Spatial thinking and STEM education: when, why, and how? *Psychology of Learning and Motivation, 57*, 147–181.

Van Gog, T., Paas, F., & Van Merrienboer, J. J. G. (2005). Uncovering expertise-related differences in troubleshooting performance: combining eye movement and concurrent verbal protocol data. *Applied Cognitive Psychology, 19*, 205–221. http://dx.doi.org/10.1002/acp.1112.

Vandenberg, S. G., & Kuse, A. R. (1978). Mental rotations, a group test of three-dimensional spatial visualization. *Perceptual and Motor Skills, 47*, 599–601. http://dx.doi.org/10.2466/pms.1978.47.2.599.

Williamson, V. M., & Jose, T. J. (2009). Using visualization techniques in chemistry teaching. In N. J. Pienta, M. M. Cooper, & T. J. Greenbowe (Eds.). *Chemists guide to effective teaching, Vol. 2*. Upper Saddle River, N.J: Prentice Hall.

Wu, H., & Shah, P. (2004). Exploring visuospatial thinking in chemistry learning. *Science Education, 88*, 465–492. http://dx.doi.org/10.1002/sce.10126.