Journal of Educational Psychology
2016, Vol. 108, No. 4, 509–527

0022-0663/16/$12.00 http://dx.doi.org/10.1037/edu0000077

Model Manipulation and Learning: Fostering Representational Competence With Virtual and Concrete Models

Andrew T. Stull and Mary Hegarty
University of California, Santa Barbara

This study investigated the development of representational competence among organic chemistry students by using 3D (concrete and virtual) models as aids for teaching students to translate between multiple 2D diagrams. In 2 experiments, students translated between different diagrams of molecules and received verbal feedback in 1 of the following 3 intervention conditions: with concrete models, with virtual models, or without models. Following the intervention, diagram translation accuracy was measured in 3 posttests, which were with models, without models, and after a 7-day delay. The virtual models in the 2 experiments differed in the level of congruence between the actions performed with the input device and the resulting movement of the virtual model. Study 1 used a low congruence interface and Study 2 used a high congruence interface. Students learned more when models were available. In terms of learning outcomes, model-based feedback was superior to verbal-feedback alone, models served as a learning scaffold rather than a crutch, and learning with model-based feedback was resilient over a 7-day delay. Finally, concrete and virtual models were equivalent in promoting learning, and action-congruence of the interface did not affect learning. The results are discussed with respect to their implications for instruction in organic chemistry and science, technology, engineering, and mathematics disciplines more generally.

Keywords: representational competence, learning scaffold, concrete models, virtual models, STEM learning

Visual representations are essential aspects of communication, research, and teaching in the science, technology, engineering, and mathematics (STEM) disciplines (Pauwels, 2006). Therefore, a crucial aspect of STEM education is learning the conventions and uses of disciplinary visual representations (Ferk, Vrtacnik, Blejec, & Gril, 2003; Trumbo, 2006), sometimes referred to as acquiring *representational competence* (Kozma & Russell, 1997). However, mastering visual representations in science is often challenging for students (Chariker, Naaz & Pani, 2011; Kali & Orion, 1996; Wu & Shah, 2004; Novick, Stull, & Catley, 2012). The focus of this study is on methods of facilitating the development of representational skills, specifically, how three dimensional (3D) models can be used to help students learn about the structure of molecules and learn the conventions of disciplinary representations in organic chemistry. We propose that external models serve as cognitive scaffolds (Yelland & Masters, 2007) for reducing cognitive load (Kalyuga, 2007; Sweller, 1988) and for externalizing and augmenting cognitive processes (Kirsh, 1995) to support the development of representational competence and spatial reasoning with visual representations.

Many STEM disciplines employ models, that is, three dimensional visuospatial representations, in teaching and practice. Models are complementary to other types of representations such as diagrams, formulas, and equations (Ainsworth, 1999). In mathematics, models, such as Cuisenaire rods, are used to physically instantiate concepts (Kaminski, Sloutsky, & Heckler, 2009) and support cognitive development from enactive toward symbolic forms of representation (Abrahamson & Lindgren, 2014). In anatomy, models typically represent the visual, spatial, and tactile properties of living tissues or organs (Preece, Williams, Lam, & Weller, 2013). In geology (Kastens & Rivet, 2010) and biochemistry (Harris et al., 2009), models represent objects and processes that occur at spatial and temporal scales that are not directly experienced. In astronomy, (Barnett, Yamagata-Lynch, Keating, Barab, & Hay, 2005) models enable students to adopt different frames of reference.

Here, we focus on organic chemistry, a domain in which spatial information is particularly important and in which spatial representations are ubiquitous. Spatial thinking is important in chemistry because the reactivity of molecules is predicted, not just by the number and type of atoms that make up a molecule, but also by their spatial configuration (Harle & Towns, 2010). Chemists use two general types of spatial representations to represent submicroscopic entities such as molecules; 3D models, which might be concrete (i.e., physical) or virtual (i.e., computer-based), and 2D diagrams, which use conventions to represent 3D relations in the

This article was published Online First October 12, 2015.
Andrew T. Stull and Mary Hegarty, Department of Psychological and Brain Sciences, University of California, Santa Barbara.
This research was supported in part by grants from the National Science Foundation (1008650 and 1252346) and the Spencer Foundation. We thank Anoush Akopyan, Trevor J. Barrett, Paula Bruice, Morgan Gainer, Soroush Kazemi, R. Daniel Little, Richard E. Mayer, David Sanosa, and Mike Stieff for their assistance, suggestions, and contributions throughout this project.
Correspondence concerning this article should be addressed to Andrew T. Stull, Department of Psychological and Brain Sciences, University of California, Santa Barbara, CA 93106. E-mail: andrew.stull@psych.ucsb.edu

two dimensions of the printed page (Francoeur, & Segal, 2004). Developing skills in drawing, interpreting, and translating between these spatial representations is essential to a student's education as a chemist (Cheng & Gilbert, 2009; Goodwin, 2008; Kozma & Russell, 1997). However, these spatial representations are not always easy to interpret and use (Harle & Towns, 2010; Keig & Rubba, 1993; Pribyl & Bodner, 1987). Specifically, beginning students have difficulty translating between different representations of the same molecule (Kozma & Russell, 1997; Wu & Shah, 2004) and this is particularly true of students with low spatial ability (Harle & Towns, 2010; Pribyl & Bodner, 1987).

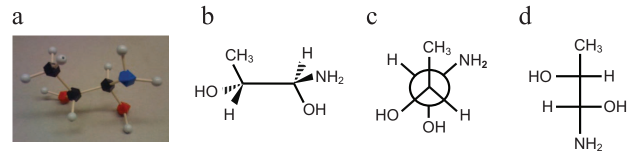


Figure 1. Four structural representations of an organic molecule. (a) A concrete (ball-and-stick) model where color is used to denote different atoms. Black is carbon, white is hydrogen, red is oxygen, and blue is nitrogen. (b) Dash-Wedge diagram (side-view), (c) Newman diagram (end view), and (d) Fischer diagram (upright view) of the same organic molecule depicted in the balland-stick model. (Originally published in Stull, et al, 2013) See the online article for the color version of this figure.

Cognitive Load and Learning Representations

Learning novel representational formalisms in organic chemistry induces high cognitive load for a number of reasons. First, the entities to be represented, molecules made up of several atoms in specific spatial configurations, are quite complex, so that mental representations are likely to overload spatial working memory (Shah & Miyake, 1996). A second source of cognitive load is that interpreting different diagrams involves recalling and imagining how different diagrammatic conventions depict 3D spatial entities in the two dimensions of the printed page (Tversky, 2005). A third source of cognitive load is that translating between different diagrammatic representations involves mentally transforming these representations. Students with poor spatial abilities are likely to experience greater cognitive load, because they have more limited working memory capacity (Shah & Miyake, 1996).

We propose that 3D models provide an external representation that allows 3D space to be represented directly, thus relieving students of the cognitive demand of interpreting how 2D diagrammatic conventions represent 3D space. With external models, 3D space is perceptually evident, offering students a holistic representation of the referent (Copolo & Hounshell, 1995; Savec, Vrtacnik, & Gilbert, 2005; Wu, Krajcik, & Soloway, 2001). Moreover, manipulating an external model allows students to also externalize the transformation processes, supporting enactive learning (Bruner, 1957; Cohen, 1989). It is easier to manually rotate an external model of a spatial entity, such as a molecule or the solar system, than it is to mentally rotate an internal representation of the referent (cf. Cary & Carlson, 2001; Kirsh, 1995; Preece et al., 2013). We propose that when a model is used to externalize the referent as well as its transformation, cognitive load is reduced for the student, and this enables the student to invest more cognitive effort in mapping conventions between the diagram and its referent and in translating between representations. Moreover, we argue that models can act as a cognitive scaffold (Yelland & Masters, 2007) to enable students to make the representational connections between diagrams and models, so that once students understand these representational connections and transformations, the models are no longer necessary and can be removed.

To illustrate this proposal, Figure 1 shows four representations of the same molecule, which are the focus of the present research. The three diagrams are 2D representations, which show the molecule from different perspectives and use different conventions (described in Appendix A) to depict three dimensional information in the two dimensions of the printed page. Interpreting these diagrams requires effortful interpretation of the spatial conventions, which must be maintained in working memory. The model is an iconic 3D representation in which the 3D relations between parts of the representation represent the 3D relations between the atoms and bonds, so that the spatial relations can be directly perceived, and there is less load on working memory.

Figure 2 outlines two possible strategies for translating from one of the diagrams (e.g., Dash-Wedge) to another (e.g., Newman). In one strategy, the *internal transformation* strategy, students view a given diagram, decode its spatial conventions to form an internal image of the molecule, mentally rotate this internal model, and then encode the spatial conventions of the new perspective before drawing the target diagram. We argue that this strategy is likely to overload working memory capacity, that is, induce high cognitive load (Kalyuga, 2007; Sweller, 1988). A second option is the *external transformation* strategy, which is to decode the conventions and manipulate a model to first align it with the perspective of the given/starting diagram (match start), then physically rotate the model to align it to the perspective of the target diagram (match target), and then encode the spatial conventions to draw the target diagram. In this way, a model can serve as a tool to support structural alignment and mapping of features between different representations (Gentner, 1983). In using this strategy, performing actions in the world rather than in the mind reduces cognitive load (cf. Cary & Carlson, 2001; Kirsh, 1995; Preece et al., 2013), freeing up mental resources to enable students to make the relevant referential connections. As a result, we might expect that students will be more successful when they use the external transformation strategy, especially if they have poor spatial abilities, which are known to be associated with limited spatial working memory (Shah & Miyake, 1996).

In previous research, Stull, Hegarty, Dixon, and Stieff (2012) demonstrated that providing students with concrete models improved performance in translating between different diagrams of molecules, but only when models were used to perform the external transformation strategy. However, most students did not spontaneously adopt this strategy, and many ignored the models. In addition, spatial ability predicted task performance in control conditions, but when models were available, model use was a much stronger predictor of translation accuracy than spatial ability. In a follow up study, Padalkar and Hegarty (2014) developed an intervention in which students first completed some diagram translation problems, and were then guided to check their solutions by trying to match a concrete model of the same molecule to both the diagram they were given and the diagram they had drawn. Completing diagram translation problems made students externalize

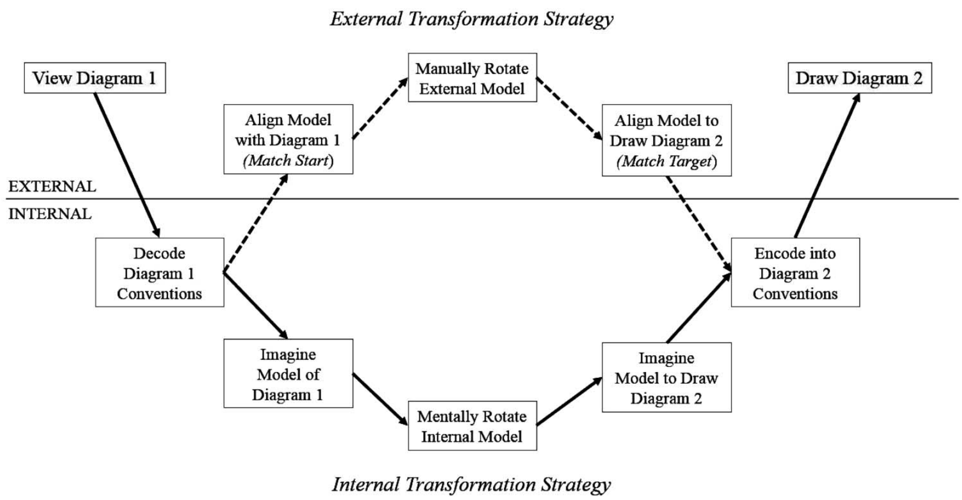


Figure 2. Process model for the translation between a viewed diagram and a diagram to be drawn. The solid arrows mark the path for the Internal Translation Strategy, which employs mental imagery, and the dashed arrows mark the path for the External Translation Strategy, which employs manual actions to enact the translation.

their existing understanding of the representational conventions. Comparing the external 3D model with their self-generated diagram exposed students to their errors and caused them to discover the benefit of models. As a result, their use of the external transformation strategy increased on later problems and their performance on the diagram translation task dramatically improved. Padalkar and Hegarty suggested that this intervention was successful because students were overconfident in their ability to complete the task without models and model-based feedback caused them to confront their illusions of understanding (Rozenblit & Keil, 2002).

In Padalkar and Hegarty's (2014) study, students were provided with concrete models on posttests and improved performance was mediated by model use, raising the concern that models might have become a crutch for students, rather than scaffolding their learning. Ultimately, students need to construct internal mental models of molecular structure and diagram translation, if only because they are not allowed to bring concrete models to their exams. Therefore, it is important to test whether effects of model-based feedback can transfer to situations in which students do not have access to models. Furthermore, Padalkar and Hegarty tested students immediately after the intervention, raising questions about the durability of the learning effects.

Scaffolds or Crutches?

The practice of scaffolding was systematically studied by Jerome Bruner (1957) in the context of children's learning, which in his situation could be referred to as social scaffolding because the support is provided through the social interaction of the child and adult tutor. Importantly, the support offered by the tutor is withdrawn as the student develops competence. In a similar way, a manipulative model may serve as a cognitive scaffold (Yelland & Masters, 2007) when it offers temporary support to help students understand how various 2D diagrams represent 3D space and how to translate between different representations. We predict that models serve as a cognitive scaffold because they relieve students of cognitive load by externally representing both the 3D structure of the object and the translation processes, and this frees up cognitive resources, enabling students to learn the conventions of 2D diagrams and how to translate between these diagrams. In the present study we test the prediction that models act as a scaffold by examining whether manipulation of concrete or virtual models during learning transfers to performance when models are no longer available. If manipulation of models during learning enables students to internalize the spatial structures and transformations, it should be possible to remove the scaffold. However, it is possible that manipulating models may make students dependent on external models. In short, we ask whether manipulative models scaffold learning or whether they act as a crutch.

There are several reasons to believe that external manipulations of models will lead to the development of internal spatial transformation processes, that is, act as a scaffold. First, basic research on enactive learning has shown that memory for verbal information can be enhanced when students enact descriptive material (Cohen, 1989; Engelkamp, Zimmer, Mohr, & Sellen, 1994; Glenberg, Gutierrez, Levin, Japuntich, & Kaschak, 2004; Schwartz & Plass, 2014). Enactive learning theorists suggest that physical action enhances multimodal coding and recall. Although previous research primarily involved learning about everyday actions (e.g., throw a ball, hammer a nail) to support reading comprehension rather than scientific processes, models might also supports memory encoding and recall by serving as cognitive scaffolds to enact mental processes. Second, there is now considerable evidence that mental and manual spatial transformations share common processes (Wohlschläger & Wohlschläger, 1998), and that performing spatial transformations manually can improve the corresponding mental spatial transformation processes. Specifically, practice in manually rotating physical or virtual objects can improve mental rotation skills (Adams, Stull, & Hegarty, 2014; Pani, Chariker, Dawson, & Johnson, 2005; Smith & Olkun, 2005; Wiedenbauer & Jansen-Osmann, 2008). We, therefore, predict that our intervention, which involves externally manipulating models to receive feedback, will not just teach students to translate diagrams by externally manipulating models, but will also enable students to internalize the relevant spatial transformations, so that learning transfers to situations in which models are no longer available.

Concrete or Virtual?

A secondary question addressed in this research is whether concrete and virtual models are equally effective as scaffolds to developing representational competence. In the chemistry classroom, concrete models are quickly taking a backseat to computer-based, virtual models which are becoming increasingly available (U.S. Government Accountability Office, 2005). However, research comparing virtual and concrete models in chemistry is sparse and in previous research, use of models was often confounded with other aspects of instruction, such as project-based learning (Barak & Dori, 2004) and case-based learning (Dori & Kaberman, 2012), making it impossible to separate the benefit of the models from the method of instruction.

In comparing concrete and virtual models in STEM domains, it is important to realize that these types of models differ in perceptual and motor cues. First, concrete models provide information by haptic and visual cues, so that a student manipulating a concrete model receives information about the model's position, shape, size, and texture from multiple modalities. In contrast, when students interact with a virtual model, the primary information is visual (Ruddle & Jones, 2001). Second, virtual models in STEM disciplines differ in their *action-congruence*, that is, correspondence between the actions performed with the interface and the resulting movement of the virtual model (Satava, 1999; Triona & Klahr, 2003). An example of high congruence is rotating an input device in three dimensions to see the corresponding 3D rotation of the on-screen object; an example of low congruence is pressing a key on a keyboard to produce the same rotation—the latter is currently more common in typical chemistry classrooms. High action-congruence has been found to be beneficial to learning manual skills, such as learning to manipulate a surgical instrument (Snyder, Vandromme, Tyra, & Hawn, 2009). In contrast, low action congruence is likely to increase cognitive load.

Basic research on mental rotation supports the idea that action-congruence may be important when using a hand-held interface to manipulate a virtual model. Wohlschläger and Wohlschläger (1998) asked people to mentally rotate objects while simultaneously rotating a nob that could be turned in the same direction or the opposite direction as the requested mental rotation. Performance was impaired when the direction of knob turning was opposite to the direction of mental rotation. Similarly, Wexler, Kosslyn, and Berthoz (1998) found that the speed of mental rotation could be increased or decreased when the speed of simultaneously rotating a joystick increases or decreases, and suggest that mental rotation is a covert simulation of manual rotation. Given the value of enactment, the common processes underlying manual and mental processes, and the impairment observed when manual and mental tasks are discordant, we predict that action-congruence will enhance performance and learning with virtual models.

The Present Study

In the present study, we examined how model use helps students learn to translate between diagrammatic representations and how this learning transfers to situations in which models are not available. We also compared learning with computer-based, virtual models and concrete models. We conducted two experiments comparing the effectiveness of model-based feedback with virtual or concrete models to verbal feedback in teaching students to translate between different diagrams of molecules. In Experiment 1, the virtual models had relatively low action-congruence. In Experiment 2 they had relatively high action-congruence.

First, we hypothesized that the model-based intervention would be associated with improved translation accuracy after instruction with models because models reduce cognitive load by allowing for internal representations and processes to be replaced or augmented by external objects and actions (Stull et al., 2012) and because model-based feedback supports knowledge integration by addressing student misconceptions (Padalkar & Hegarty, 2014).

Second, we hypothesized that models scaffold learning because they relieve cognitive load by representing 3D spatial relations directly and allowing students to partly externalize spatial transformation processes, freeing up mental resources for learning. Specifically, students who learn with models should have better performance than those who learn without models, even when models are no longer available.

Third, we hypothesized that students who learned with models would show learning gains that are resilient after a delay, even in the absence of models, because external transformations enacted with the manipulative models lead to internalization of diagram conventions and translation processes.

Fourth, we hypothesized that using models to externalize key aspects of the representation translation would be more predictive of performance than spatial or reasoning ability, as found by Stull et al. (2012). Moreover, we predicted that using models in this way would predict performance not just when models are available, but would predict future performance when models are no longer available because manipulative models directly represent 3D space and can be used to enact the translation process, limiting demands placed on spatial working memory if the object and process were only imagined.

Finally, although we expected learning with both virtual and concrete models to be better than learning without models, we predict superior learning for students who used concrete models compared to those who use low-congruence virtual models, because low action congruence adds cognitive load, thus reducing cognitive resources available for learning.

Study 1

Method

Participants and design. Participants were 148 undergraduate organic chemistry students at a research university who had completed at least one course in organic chemistry and had been introduced to the models and diagrammatic representations in their courses. One hundred and five students remained in the analysis (age: $M = 19.6$, $SD = 0.99$) after 4 were dropped for failing to follow directions, 4 because of technical errors, 11 for not completing the study, and 24 because they had perfect or almost perfect performance on the pretest (i.e., they made either zero or one error), indicating that they had already mastered the task and there was nothing for them to learn.

The experiment followed a 4 (test type: pretest, models posttest, no-models posttest, delayed posttest) × 3 (intervention: concrete models, virtual models, or control) mixed design. Test type was a within-subjects variable and intervention was a between-subjects variable. Males and females were assigned randomly to one of the two models groups or the control group, to ensure gender balance across the conditions. There were 36 (17 males) in the concrete model group, 31 (15 males) in the virtual model group, and 38 students (16 males) in the control group. They received either course credit or $40 for their participation.

Materials. The study materials included an informed consent sheet, a diagram description sheet, a task description sheet, diagram translation problems (one 6-item pretest and three 12-item posttests), ball-and-stick models (12 concrete and 12 virtual models), a demographic questionnaire, a spatial ability test, and an abstract reasoning test.

Diagram translation problems. For each problem, one kind of diagram of a molecule (i.e., either Dash-wedge, Newman projection, or Fischer projection; see Figure 1) was given at the top of a paper worksheet (8.5 in. × 11 in.) with text appearing below giving instructions to translate this diagram into one of the other two kinds of diagram of the same molecule. A horizontal line on the worksheets for the pretest divided the paper into an upper region where participants drew their original translation and a lower region where additional drawings were made if the original was incorrect. The drawing space was not divided on the posttests. There were six kinds of problems (i.e., translations from Dash-wedge to Newman, Newman to Fischer, Fischer to Dash-wedge, and vice versa).

The pretest included six unique problems (i.e., one of each kind) with 4-carbon molecules. The three separate posttests (models, no-models, and delayed) each included 12 unique problems, six with 4-carbon molecules and six with 5-carbon molecules. All of the molecules used in the problems had two chiral carbons (i.e., carbons that were bonded to four different atoms or groups of atoms) linked by a single bond. The problems for the no-models posttest were diastereomers (i.e., differed from the pretest problems in the order of components around one chiral carbon) and problems for the delayed posttest were enantiomers (i.e., differed from the pretest problems in the order of components around both chiral carbons).

Ball-and-stick models. The concrete models were constructed from a commercial molecular modeling kit (HGS Introductory Organic Chemistry Set 1000) that is commonly used in high school and college chemistry courses. The virtual models were created with ACD/ChemSketch© (Toronto, Canada) and Blender© (Amsterdam, The Netherlands). Visual appearance of the virtual models matched that of the concrete models in apparent size, shape, and color to minimize differences between the concrete and virtual stimuli. Vizard© (Santa Barbara, U.S.A.) virtual reality software was used to display the virtual models. The virtual models were manipulated with a mouse and keyboard in a manner that is typical of most commercially available molecular modeling software that is available to students in a high-school or college chemistry course. The whole virtual model could be rotated with the computer mouse. A rotation or twist around the central bond linking the two chiral carbon atoms could be performed with the up- and down-keys on the keyboard. The bonds connecting the substituents to the chiral carbon atoms could be rotated in the concrete models but not in the virtual models, which constrains the interactivity.

Questionnaire and tests. The demographic questionnaire, spatial ability test, and abstract reasoning test, were administered with the Qualtrics© (Utah, U.S.A.) online survey tool. The questionnaire collected self-reported demographic information (i.e., age, sex, date of birth, major course of study, years in college, handedness, colorblindness, and stereo vision) as well as specific information about how many organic chemistry classes students had taken and their familiarity with video game technology, such as 3D glasses or TVs, and with the diagrams and models used in the study. An online version of the Vandenberg and Kuse (1978) Mental Rotation Test was administered as a test of spatial ability, which consisted of 20 items administered in two 3-min blocks of 10 items. An online version of the Abstract Reasoning Test from the Differential Aptitudes Test (Bennett, Seashore, & Wesman, 1974) was administered as a test of general reasoning ability (40 items, 10 min time-limit).

Procedure. The experimental and control groups were first given basic instructions, which included examples of the three kinds of diagrams with reminders of the conventions of each diagram (see Appendix A) and the nature of the task (see Appendix B). After they read the instructions, participants completed the Pretest problems without the aid of models. They were told that they could draw any conformation of the molecule shown in the given diagram.

After completing the pretest problems, all groups went through a training intervention using their pretest drawings (described in the following paragraphs). Next, all groups solved the 12-problem posttest, with models available (models posttest) for the concrete and virtual groups, then they completed an additional set of 12 problems with no models available (no-models posttest), and responded to the demographic questionnaire and the Mental Rotation Test (MRT). Then each participant was scheduled to return in 7 to 10 days to complete the second session.

In the second session, participants completed the 12-item delayed posttest and the Abstract Reasoning Test before being debriefed and dismissed. Participants were videotaped with their consent during the Pretest and all posttest drawing tasks.

Intervention. The intervention involved providing participants with feedback to help them check their solutions to each of the Pretest problems and to redraw a new solution if any of their drawings were incorrect. In the case of the Control group, only verbal feedback was provided but for the two model groups, verbal as well as model-based feedback was provided.[1] All three groups (concrete models, virtual models, no-models) were allowed to refer to the instruction sheet describing the three types of diagrams during the intervention. Table 1 summarizes the similarities and differences between the different intervention conditions.

Participants in the control group were first instructed to review the given diagram and to make sure that they understood its conventions. Next, they were instructed to review their drawn diagram and to tell the experimenter if they thought their diagram

[1] This design differs from the study of Padalkar and Hegarty (2014; Experiment 1) in which the control group received no feedback.

Table 1
Summary of Feedback Type (Concrete Model, Virtual Model, or Control) and Timing of Tests for Both Study 1 and Study 2

Feedback type	Pretest	Intervention	Models posttest (same day)	No-models posttest (same day)	Delayed posttest (7-day later)
Concrete	No models	Concrete models with verbal feedback	Concrete models	No models	No models
Virtual	No models	Virtual models with verbal feedback	Virtual models	No models	No models
Control	No models	Verbal feedback only	No models	No models	No models

was correct or incorrect. Students were then told if their solution was correct or incorrect and the nature of any error was described (see the following text). Finally, students drew a new diagram below the horizontal line, and received verbal feedback on that diagram. These steps were repeated until each participant drew a correct solution to the problem. The number of corrective intervention cycles to reach criterion, a correct answer, on each Pretest problem were tabulated.

For the model-based feedback for the two model groups, participants were provided with the appropriate model of the molecule that they had attempted to draw in each of the six Pretest problems. First, they were asked to match (i.e., structurally align, cf., Gentner, 1983) the model with the given diagram in the problem. Guidance was provided only if the participant was unable to align the model to the given diagram. Next, participants were asked to align the model with their solution to the problem, which was drawn above the horizontal line, and to determine if their diagram was correct. If the participant had drawn a correct solution, alignment was possible. Once they could align the model with their drawing, they were asked to move to the next problem. If the solution was incorrect, it was not possible to structurally align the model with the solution, and this was explained if the participant was not able to discover his or her error. Then participants were asked to use the model to draw a new, corrected diagram below the horizontal line. If necessary, these steps were repeated until each participant drew a correct solution to the problem, with the number of cycles tabulated for each problem.

The verbal feedback provided during all interventions was error-specific. The recorded errors were either spatial errors, connectivity errors, or fundamental errors (Padalkar & Hegarty, 2014). A spatial error occurred when the drawn diagram was made up of the correct molecular substituents connected to the correct chiral carbon atoms, but their 3D spatial arrangement was incorrect. A connectivity error occurred where the drawn diagram was made up of the correct substituents, but these substituents were connected to the wrong chiral carbon atoms. Finally, a fundamental error was syntactic and including drawing the wrong type of diagram, drawing the wrong template for a diagram type, or drawing a diagram with missing, additional, or duplicate substituents. An example of verbal feedback for a spatial error is: "You have drawn the wrong order of these three substituents, so you have drawn an isomer and not the molecule requested." An example of verbal feedback for connectivity errors is "In your diagram you have drawn these substituents on the wrong carbon." Verbal feedback for a fundamental error depended on the nature of the error. For example, if the student drew the wrong diagram template then they might be told "Notice that you were asked to draw a Newman projection, but you drew a Fischer," or, if a participant drew a duplicate substituent, they were told, "Notice that there is only one group in the given diagram. You have drawn two."

Out of 641 pretest trials, 205 (32%) were correct on their first drawing, 379 (59%) were corrected in the first cycle of the intervention, 51 (8%) were corrected in the second cycle of the intervention, and there were 6 cases (~1%) in which participants took three or more attempts to draw a correct solution.[2]

Coding of solutions. Each diagram translation problem was coded as correct or incorrect (no partial credit was given). The total proportion of correct solutions served as the accuracy scores for the pretest and posttest measures. Pretest and posttest data for 25 participants (23% of the data; 1050 trials) were coded independently by two researchers and interrater reliability was high (Cohen's $\kappa = 0.93$). Discrepancies were resolved by consensus.

Coding of model use behaviors. Model use was coded from the videos by two experienced coders for three behaviors central to the external transformation strategy, namely (1) aligning the model to match the conformation and the orientation of the given or starting diagram (match-start), (2) reconfiguring the model by rotating substituents around the sigma bonds (rotate bond), and (3) aligning the model to match the conformation and orientation of the model to the target diagram (match-target). Video recordings of 15 participants (21% of the data; 840 trials) were coded independently by two reviewers and interrater reliability was high (Cohen's $\kappa = 0.93$). Discrepancies were resolved by consensus.

Results and Discussion

Throughout this article, nonparametric tests were performed if assumptions of analyses of variance were not met and, when appropriate, Bonferroni corrections were applied to address family wise error rate. There were no statistically significant differences between the experiment groups in age, mental rotation ability, abstract reasoning, grade point average, or number of organic chemistry courses completed. The relevant descriptive and inferential statistics are given in Table 2. There was a statistically significant difference between the groups in the number of intervention cycles needed to reach criterion on pretest problems. Participants in the control group required more corrective cycles than did those in the concrete group, $U = 454.5$, $Z = -2.52$, $p < .01$, $r = .29$,[3] but not than those in the virtual group, $U = 437.5$, $Z = -1.85$, $p = .03$. The concrete and virtual groups did not statistically differ, $U = 518.5$, $Z = -0.51$, $p = .31$.

Models as learning aids. Figure 3a and Table 3 present the pretest and posttest accuracy data for the different groups. As revealed by a Kruskal-Wallis test, participants in the three intervention conditions did not statistically differ in pretest perfor-

[2] Spatial errors were the most common, and accounted for 94% of all errors in Study 1. The percentage of spatial errors did not differ across conditions.

[3] Effect size measures, Pearson's r, are calculated for non-parametric statistical comparisons as suggested by Field (2005). Rosnow, Rosenthal, and Rubin (2000) outline methods for converting between Pearson's r and Cohen's d.

Table 2
Means, Standard Deviations, and Inferential Statistics Comparing Individual Difference Measures for the Three Intervention Conditions in Study 1

Measure	Condition	*M* (*SD*)	*n*	Inferential test
Age	Concrete	19.46 (.80)	36	
	Virtual	19.49 (.66)	31	$H(2) = 2.31, p = .32$
	Control	19.90 (1.29)	38	
MRT	Concrete	40.49 (14.07)	36	
	Virtual	42.91 (16.31)	31	$H(2) = 2.02, p = .37$
	Control	36.97 (17.89)	38	
Abstract reasoning	Concrete	25.17 (6.37)	36	
	Virtual	25.46 (6.65)	31	$H(2) = .40, p = .82$
	Control	25.84 (6.95)	38	
GPA	Concrete	3.24 (.45)	36	
	Virtual	3.39 (.39)	31	$H(2) = 3.62, p = .16$
	Control	3.21 (.33)	37	
Number of O-chem courses	Concrete	1.89 (.74)	36	
	Virtual	1.74 (.56)	31	$H(2) = 4.5, p = .13$
	Control	2.08 (.66)	38	
Number of intervention cycles	Concrete	4.33 (1.60)	36	
	Virtual	4.58 (1.65)	31	$H(2) = 6.9, p = .03$
	Control	5.37 (1.78)	38	

Note. Kruskal-Wallis Tests were performed because assumptions of analyses of variance were not met for some measures. MRT = Mental Rotation Test; GPA = grade point average.

mance, $H(2) = 0.34$, $p = .84$, but they did differ in performance on the models posttest, $H(2) = 48.25$, $p < .001$, no-models posttest, $H(2) = 12.01$, $p < .01$, and delayed posttest, $H(2) = 7.32$, $p = .02$.

A series of paired contrasts were conducted to further investigate these results. First, we investigated the immediate and delayed benefits of receiving model-based feedback with either concrete or virtual models to only verbal feedback to test the hypothesis that model based feedback is more effective. Next, we compare the concrete and virtual model conditions. Finally, we examined which model-use behaviors predict current and future performance.

Models versus no models. To investigate the benefit of receiving models, performance of students in the concrete and virtual model conditions were combined and compared to students in the control group on the three posttest measures. We predicted that participants in the models groups would outperform the control group on the posttest with models available. If external manipulations of models lead to internalization of the spatial transformation process, participants in the models group would also be more accurate when models are no longer available and after a delay.

Does the model-training intervention support diagram translation? A Mann–Whitney test for the models posttest revealed that translation accuracy was statistically higher for the models group over the control group, $U = 253.0$, $Z = -6.92$, $p < .001$, $r = .67$, replicating earlier work (Padalkar & Hegarty, 2014) and providing evidence that models support diagram translation accuracy.

Do models act as a scaffold or as a crutch? A Mann–Whitney test for the no-models posttest revealed that translation accuracy was statistically better for the models group than the control group, $U = 758.0$, $Z = -3.45$, $p < .001$, $r = .34$, suggesting that models acted as a scaffold to learning rather than a crutch.

Is learning with models resilient over a delay? A Mann–Whitney test for the delayed posttest revealed that translation accuracy was statistically higher for the models groups than for the control group, $U = 873.5$, $Z = -2.68$, $p = .004$, $r = .26$. These results demonstrate that the model-based intervention supported retention of what was learned when using the models.

As shown in Table 4, performance on the delayed posttest was statistically better than on the pretest for both the models group and the control group, indicating that both groups learned from their respective interventions, although improvement was statistically greater for the models group than for the control group, as reported earlier. For the models group, performance at the delayed posttest was statistically poorer than at the models posttest, indicating that some support offered by the models was not retained over the delay. The control group was not statistically different between the delayed posttest and the models posttest. Finally, performance was not statistically different in the no-models posttest, which occurred immediately after the intervention, and the delayed posttest, which occurred a week later, for either the models or control groups, indicating that performance levels remained stable over the delay.

Comparison of concrete and virtual models. We predicted less learning with virtual models, because the low action-congruence of the virtual model interface would add cognitive load, and this in turn would impede learning.

Do virtual models require more time than concrete models? Median times for completing the set of 12 problems in the concrete and virtual models conditions were analyzed with a Mann–Whitney test to examine performance effort. The analysis revealed a statistical difference between the two model conditions, $U = 156.5$, $Z = -4.58$, $p < .001$, $r = -0.58$. As shown in Figure 4, students took more time to solve the problems with the virtual models than with the concrete models during the models posttest,

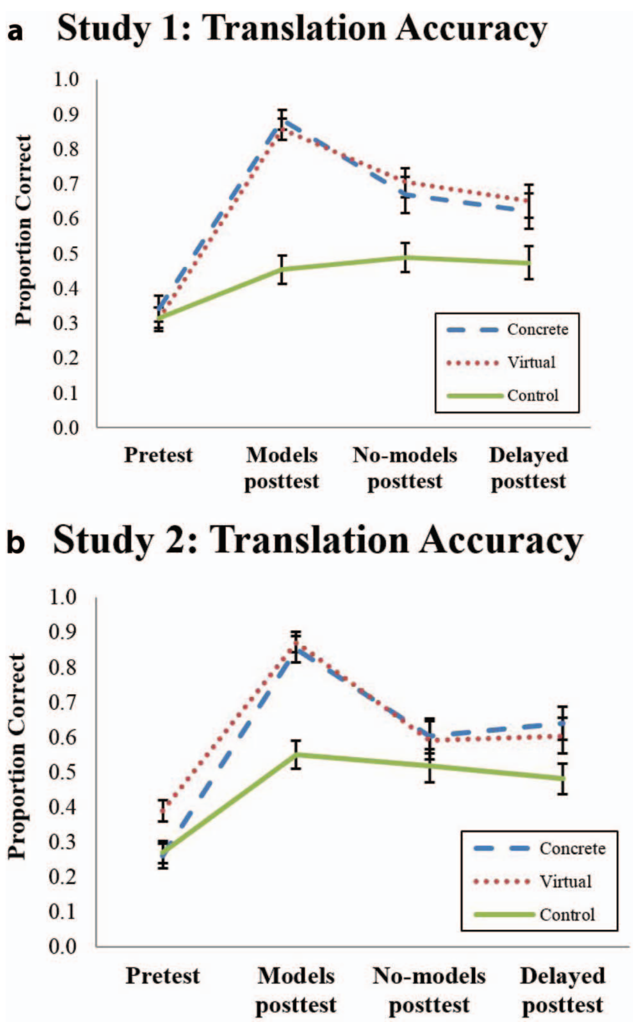


Figure 3. Graph of mean learning performance for students receiving concrete and virtual model-based interventions and a control group over measures of no-model and delayed testing. Models were only available during the pretest and intervention phase for the two model groups. Error bars indicate standard error. See the online article for the color version of this figure.

time per problem generally decreased over time, and the difference between concrete and virtual models narrowed, suggesting that students adapted to the virtual model interface over time.[4] These results are consistent with the assumption that virtual models impose more cognitive load than concrete models, at least initially.

Are learning effects greater with concrete models than with virtual models? Mann–Whitney tests comparing the concrete and virtual models groups at each of the three posttests revealed no statistically significant differences between the two models groups on the models posttest, $U = 488.5$, $Z = -0.92$, $p = .18$, the no-models posttest, $U = 549.5$, $Z = -0.11$, $p = .46$, or the delayed postest, $U = 534.5$, $Z = -0.30$, $p = .38$. These results are contrary to our predictions. In sum, students using the virtual models took more time to complete the problems but achieved similar learning outcomes to those who used concrete models.

Predictors of learning with models. We predicted that using models to externalize key aspects of the translation, such as matching the model to the desired orientation of the target diagram (see Figure 2), would predict future performance when models are not available. Correlations among three common model-use behaviors (i.e., *match-start*: aligning the model to the given diagram; *rotate bond*: performing an internal rotation of the model to match the conformation of the molecule; and *match-target*: aligning the model to the diagram to be drawn) and individual difference measures (i.e., mental rotation and abstract reasoning) are reported in Tables 5 and 6.

What predicts performance when models are available? A stepwise multiple regression analysis was conducted to examine how measures of spatial and reasoning abilities and use of models predicted translation accuracy with the aid of models (see the models posttest results on the left portion of Table 7). Entered in the first step, the two ability measures explained a nonstatistically significant 7% of the variance in translation accuracy (multiple $R = .26$). Entered in the second step, the three model use behaviors explained an additional and statistically significant 52% portion of the variance. Consistent with previous research (Stull et al., 2012), the partial regression coefficients revealed that match target ($sr^2 = .37$) was the only statistically significant and unique predictor of translation accuracy after controlling for the other variables.

What predicts performance when models are no longer available? To assess predictors of diagram translation performance in the absence of models (no-models posttest), a second stepwise multiple regression analysis was conducted (See middle portion of Table 7). In the first step, the two ability measures explained a statistically significant 13% portion of variance (multiple $R = 0.36$) in diagram translation accuracy. Examination of the partial regression coefficients revealed that neither mental rotation performance nor abstract reasoning alone explained a statistically significant portion of the variance in the first step. In the second step, the three model use behaviors explained an additional and statistically significant 17% of the variance in translation accuracy. Of the partial regression coefficients, match target was the only statistically significant predictor of translation accuracy ($sr^2 = 0.12$) after controlling for the other variables. These results are novel in that they show that how people use models predicts performance when models are no longer available, and are consistent with the view that external uses of models when they are available leads to the internalization of the processes of translating between representations.

What predicts performance after a delay? To assess predictors of diagram translation performance in the delayed posttest, a third stepwise multiple regression analysis was conducted (see the right portion of Table 7). In the first step, the ability measures explained a statistically significant 14% of the variance in accuracy (multiple $R = .37$). In the second step, three model use behaviors explained an additional and statistically significant 12% of the variance. Examination of the partial regression coefficients revealed that mental rotation performance ($sr^2 = 0.06$) was a statistically significant predictor of translation accuracy in the first step. In the second step, match

[4] A second analysis also showed that students in the control group did not perform statistically differently from those in the concrete models group.

Table 3
Medians, Means, and Standard Error Measures for Diagram Translation Accuracy for Study 1 and Study 2

Study	Condition	Pretest *Mdn M (SE)*	Models posttest *Mdn M (SE)*	No-models posttest *Mdn M (SE)*	Delayed posttest *Mdn M (SE)*	*n*
Study 1	Concrete	.333 .343 (.037)	1.00 .884 (.028)	.750 .669 (.052)	.750 .623 (.051)	36
	Virtual	.333 .312 (.034)	.917 .858 (.031)	.750 .704 (.042)	.667 .651 (.048)	31
	Models	.333 .328 (.025)	.917 .872 (.020)	.750 .685 (.034)	.750 .636 (.035)	67
	Control	.333 .316 (.029)	.417 .454 (.041)	.500 .489 (.042)	.458 .474 (.047)	38
Study 2	Concrete	.333 .260 (.035)	.917 .850 (.038)	.625 .603 (.051)	.625 .640 (.047)	34
	Virtual	.333 .389 (.030)	.917 .871 (.029)	.583 .591 (.055)	.500 .604 (.051)	33
	Models	.333 .323 (.025)	.917 .861 (.024)	.583 .597 (.037)	.583 .622 (.035)	67
	Control	.167 .270 (.032)	.583 .550 (.040)	.417 .518 (.048)	.417 .480 (.044)	37

Note. The models group combined both the concrete and the virtual models groups.

target ($sr^2 = 0.08$) was the only statistically significant predictor of accuracy, after controlling for mental rotation performance spatial ability and the other variables. These results suggest that how one interacts with the models, and not just the presence of a model, affects learning. They also suggest that individual differences in spatial skills become more predictive when models are not available.

In summary, the results of Study 1 support the idea that the model-based intervention scaffolds learning so that performance gains from models are seen even when models are not available and after a delay. For example, students who received the model-based feedback intervention were about 64% accurate on the delayed posttest whereas those who received only verbal feedback intervention were only about 45% accurate on the same test. Also, it is notable that students learned more in the models conditions, even though they went through fewer feedback cycles to meet criterion performance in the intervention. The results also indicate that the concrete and virtual models used here do not differ in their support of learning. Finally, students learn more from models, and have more lasting gains when they use the model to physically enact the representation translation by using the model to represent the desired orientation and conformation of the target molecule.

Study 2

The virtual models used in Study 1 used a mouse and keyboard interface with low action-congruence. The equivalent learning gains for concrete and virtual models in that experiment suggested that information gained from touching, holding, and moving the concrete models (i.e., shape, size, configuration, etc.) does not enhance learning, that is, the visual information provided by the model was sufficient for learning. In recent research, use of high-congruence virtual models in a molecule manipulation task resulted in decreased time to completion compared to concrete models (Barrett, Stull, Hsu, & Hegarty, 2014; Stull, Barrett, & Hegarty, 2013), which the authors attribute a virtual model interface that constrained users to task-appropriate actions. It seems plausible that a direct manipulation interface with high action-congruence and task-appropriate constraints might support learning even more than concrete models. We examined this possibility in Study 2.

Although otherwise identical to Study 1, Study 2 compared virtual models with high action-congruence to concrete models. We hypothesized that learning with these virtual models would be greater than learning with concrete models. On the basis of the results of Study 1, we also hypothesized that the model-based intervention would be associated with improved translation accuracy compared with the control condition with verbal feedback, that performance gains would be evident when models were no longer available and that learning gains due to the model-based intervention would be resilient after a 7-day delay even in the absence of models. Finally, we expected that interactions with the model, such as matching the model to the intended diagram, would be a predictor of future performance without models.

Table 4
Inferential Statistics Comparing Long-Term Learning as Measured at the Delayed Posttest Compared With Prior Performance at Pretest, Models Posttest, and No-Models Posttest for Studies 1 and 2

		Pretest			Models posttest			No-models posttest		
Study	Condition	*Z*	(*p*)	*r*	*Z*	(*p*)	*r*	*Z*	(*p*)	*r*
Study 1	Model	−5.75	(<.001)	−.70	5.94	(<.001)	−.73	2.05	(.04)	−.25
	Control	−3.35	(.001)	−.54	−.68	(.50)	−.11	−.37	(.71)	−.06
Study 2	Model	−5.77	(<.001)	−.70	6.40	(<.001)	−.78	−.89	(.37)	−.11
	Control	−3.79	(<.001)	−.62	2.45	(.01)	−.40	1.27	(.20)	−.21

Note. Wilcoxon's signed-rank test was used to compare groups and a Bonferroni correction was applied to set the alpha level at .017 to address family-wise error rates. The sign of the *Z* value indicates the direction of the effect compared with delayed posttest performance.

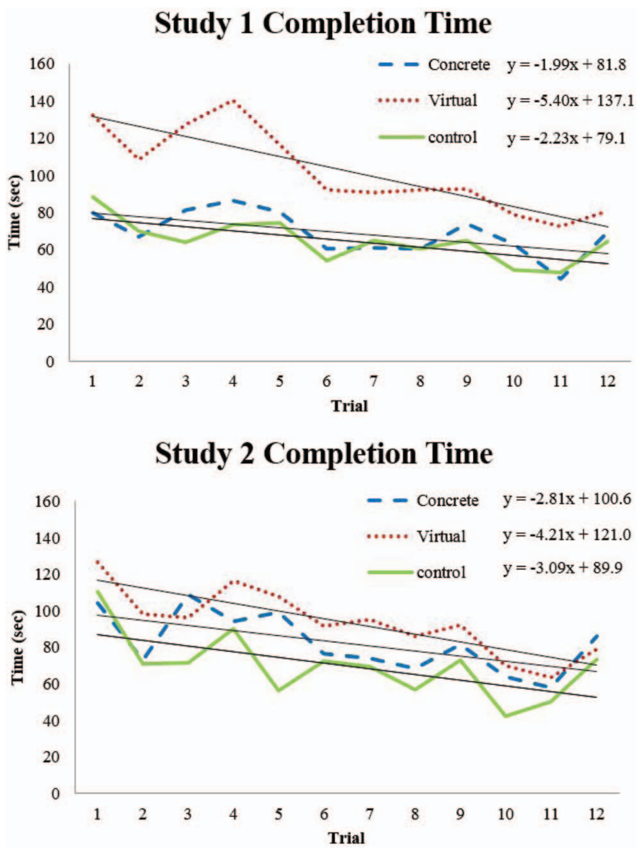


Figure 4. Mean times for completing each of 12 trials in the three conditions for Study 1 and 2. See the online article for the color version of this figure.

Method

Participants and design. Participants were 119 undergraduate organic chemistry students at a research university who completed at least one course in organic chemistry and had been introduced to the diagrammatic representations in their lectures and textbooks. One hundred and four students remained (age: M = 19.6, SD = 0.99) after 1 was dropped for failing to follow directions, 7 for not completing the study, and 7 because they had perfect or almost perfect performance (i.e., they made either zero or one error) on the pretest, which indicated that they had already mastered the task and there was nothing for them to learn. As in Study 1, participants were alternately assigned to one of the two models groups or the control group in order to achieve gender balance. Thirty-four students (13 male), were in the concrete model Group 33 (11 male) were in the virtual model group and 37 (12 male) were in the control group. Participants received course credit and $40 for their participation.

Materials and apparatus. The study materials were identical to those used in Study 1 except that the virtual models were delivered in a perceptually controlled virtual reality modeling system (Stull et al., 2013). This system was modeled after an integrated graphic and haptic system developed by Ernst and Banks (2002), which was configured to portray the illusion that a displayed virtual model was directly manipulated with the participants' colocated hands via a hand-held interface (see Figure 5). The hand-held interface was designed to allow both global rotation of the whole virtual object and local rotations of the two halves of the model around the central carbon–carbon bond, as both of these types of rotation are necessary for the tasks presented in this paper. In addition, stereo glasses were used to provide stereoscopic depth cues, enhancing the illusion that the participants were using real rather than virtual models.

Table 5
Correlation Between Predictors and Diagram Translation Accuracy for Study 1 and 2

Predictors	Pretest r	Models posttest r	No-models posttest r	Delayed posttest r	n
Study 1					
Match target	.099	.748***	.519***	.448***	69
Rotate bond	.155	.373**	.285*	.182	69
Match start	.013	.375**	.208	.203	69
Mental rotation	.232	.242*	.301*	.368**	69
Abstract reasoning	.149	.177	.301*	.312**	69
Study 2					
Match target	−.013	.708***	.330**	.422**	69
Rotate bond	−.155	.378***	.061	.212	69
Match start	−.146	.249*	.191	.203	69
Mental rotation	.045	.235*	.357**	.377**	70
Abstract reasoning	.102	.057	.159	.190	70

Note. Video data were not collected for three of the participants in Study 1, which precluded tabulating model-use behaviors. Cronbach's alpha was highly reliable for the 20-item mental rotation measure (Study 1: α = .77; Study 2: α = .76) and the 40-item abstract reasoning measure (Study 1: α = .94; Study 2: α = .90).
* p < .05. ** p < .01. *** p < .001.

Procedure. The procedures for Study 2 were identical to that for Study 1. Out of 641 pretest trials, 205 (32%) were correct, 379 (59%) were corrected in the first cycle of the intervention, 51 (8%) were corrected in the second cycle, and there were 6 cases (approximately 1%) in which participants took three or more attempts to draw a correct solution.[5]

Results and Discussion

There were no statistically significant differences between the groups in age, mental rotation ability, abstract reasoning ability, grade point average, or number of organic chemistry courses completed. There was a statistically significant difference between the groups in the number of intervention cycles needed to reach criterion on Pretest problems. Relevant descriptive and inferential statistics are given in Table 8.

Participants in the control group required statistically significantly more corrective cycles than did those in the virtual models group, $U = 354.0$, $Z = -3.07$, $p < .01$, $r = .37$, but not compared to those in the concrete models group, $U = 555.0$, $Z = -0.87$, $p = .19$. The Concrete and Virtual groups were not statistically significantly different, $U = 374.5$, $Z = -2.40$, $p = .02$, $r = .29$.

Models as learning aids. We predicted that participants who received models would perform better on the models posttest,

[5] Spatial errors were the most common, and accounted for 90% of all errors in Study 2. The percentage of spatial errors did not differ across conditions.

Table 6
Intercorrelations Between Predictors for Study 1 and 2

Predictors	Rotate bond *r*	Match start *r*	Mental rotation *r*	Abstract reasoning *r*
Study 1				
Match target	.520**	.378**	.324**	.261*
Rotate bond		.352**	.246	.236
Match start			.070	.071
Mental rotation				.327**
Study 2				
Match target	.641**	.396**	.127	.159
Rotate bond		.651**	.061	−.009
Match start			.113	.036
Mental rotation				.391**

* $p < .05$. ** $p < .01$.

no-models posttest, and delayed posttest, as in Experiment 1. We also predicted that learning with high-congruence virtual models would be greater than with concrete models. Finally, we predicted that structurally aligning the given models to the desired orientation of the target diagrams (i.e., match-target) would predict future performance when models are not available.

Figure 3b and Table 3 present the accuracy data for the pretest and three posttests for Study 2. There was a statistically significant difference between the three conditions on pretest performance, $H(2) = 9.49$, $p = .01$. Post hoc comparisons revealed that the virtual models group performed significantly better than the control group, $U = 385.0$, $Z = -2.75$, $p < .01$, $r = .33$, and the concrete model group, $U = 359.5$, $Z = -2.62$, $p = .01$, $r = .32$.[6] As in Experiment 1, we first investigated the immediate and delayed benefits of receiving model-based feedback versus receiving only verbal feedback. Next, we investigated the comparison of high action-congruence virtual models with concrete models. Finally, we investigated the effects of abilities and use of models on performance and learning outcomes.

Models or no models. To investigate the benefit of receiving model-based feedback, the two model groups were combined and compared with the control group on the three posttests.

Does the model-training intervention support diagram translation? A Mann–Whitney test for the models posttest revealed that translation accuracy was statistically higher for the models group than for the control group, $U = 415.50$, $Z = -5.69$, $p < .001$, $r = .57$. These results replicate Study 1 and earlier work (Padalkar & Hegarty, 2014; Stull et al., 2012) and indicate that models support diagram translation.

Do models act as a scaffolding for learning or as a crutch? A Mann–Whitney test for the no-models posttest showed no statistically significant difference between the model and control groups, $U = 1047.0$, $Z = -1.31$, $p = .09$. This result does not support our prediction and does not replicate the result of Study 1, which showed the predicted effect.

Is learning with models resilient over a delay? A Mann–Whitney test for the delayed posttest revealed that the model group was statistically more accurate than the control group, $U = 883.0$, $Z = -2.44$, $p < .01$, $r = .24$. That is, although there was no statistically significant immediate benefit of the model interventions, the predicted benefit was evident in longer term performance, as in Study 1.

As with Study 1, performance on the delayed posttest was also compared with performance on the pretest, the models posttest, and the no-models posttest. Results replicated those for the models condition from Study 1 but differ in showing that students in the control group showed a statistically significant decrease in performance between models posttest and the delayed posttest. These analyses are summarized in Table 4.

Comparison of concrete and virtual models. We predicted that performance would be better with high-congruence virtual models than with concrete models because high-congruence virtual models impose less cognitive load because they mimic important perceptual features of the concrete models and impose interactive constraints to limit the users' interactions with the model to task-relevant interactions.

Do interactions with high-congruence virtual models take more time than interactions with concrete models? Mean times for completing the set of 12 problems with concrete and virtual models did not differ, based on a Mann–Whitney test, $U = 509.5$, $Z = -0.65$, $p = .52$, $r = -0.08$.[7] Trial times for Study 2 are plotted on the bottom graph in Figure 4. These results show no evidence that the virtual models impose more cognitive load.

Is learning better with high-congruence virtual models than with concrete models? Mann–Whitney tests were used to compare the concrete models group with the virtual models group at each of the three posttests. There was no statistically significant difference between the two model groups on the Models Posttest, $U = 558.0$, $Z = -0.04$, $p = .48$, the no-models posttest, $U = 553.5$, $Z = -0.10$, $p = .46$, or the delayed posttest, $U = 520.0$, $Z = -0.52$, $p = .30$. The equivalence between these two groups is contrary to our predictions.

Predictors of learning with models. Correlations between the accuracy measures, the three model-use behaviors (i.e., match-start, rotate bond, and match-target), and ability measures (i.e., mental rotation and abstract reasoning) are reported in Tables 5 and 6.

What predicts performance when models are available? A stepwise multiple regression analysis was conducted (See left portion of Table 9) to examine predictors of translation accuracy on the models posttest. The two ability measures explained a nonstatistically significant 4% of the variance in translation accuracy ($R = .19$) in the first step. The three model use behaviors explained an additional and statistically significant 48% portion of the variance in the second step. The partial regression coefficients revealed that match target ($sr^2 = 0.32$) was the only statistically significant predictor of translation accuracy after controlling for the other variables.

What predicts performance when models are not available? A second stepwise multiple regression analysis was conducted (see Table 9) to assess diagram translation performance in the absence of models (no-models posttest). In the first step, the two individual

[6] There were more low performing participants in the control and concrete models groups. In a supplementary analysis, we omitted the lowest performing students for all three groups, which equalized pretest performance and revealed that the main results regarding the posttest did not change meaningfully.

[7] A second analysis also showed that differences in performance of students in the control group and the concrete models group were not statistically different.

Table 7
Hierarchical Multiple Regression Analyses Predicting Diagram Translation Accuracy From Individual Difference Measures (Step 1) and Model-Use Behaviors (Step 2) for Study 1

Predictors	Models posttest			No-models posttest			Delayed posttest		
	B	*SE*	β	*B*	*SE*	β	*B*	*SE*	β
Step 1									
Constant	.71	.09		.28	.14		.23	.15	
Mental rotation	<.01	<.01	.20	<.01	<.01	.22	.01	<.01	.26*
Abstract reasoning	<.01	<.01	.11	.01	.01	.22	.01	.01	.19
R^2		.07			.13			.14	
F		2.21			4.66*			4.77*	
Step 2									
Constant	.25	.12		−.13	.25		.06	.27	
Mental rotation	<–.01	<.01	−.01	<.01	<.01	.09	<.01	<.01	.18
Abstract reasoning	<.01	<.01	−.02	<.01	<.01	.15	.01	.01	.14
Match start	<.01	<.01	.06	<.01	.01	.04	.01	.01	.11
Rotate bond	<−.01	.01	−.02	−.01	.03	−.02	−.03	.03	−.15
Match target	.06	.01	.76***	.06	.02	.44**	.05	.02	.37*
R^2		.59			.31			.25	
ΔR^2		.52			.17			.12	
$F\Delta R^2$		24.69***			4.80**			3.08*	

Note. $N = 64$. Data conformed to statistical assumptions of the regression analysis.
* $p < .05$. ** $p < .01$. *** $p < .001$.

difference measures explained a statistically significant 11% of variance ($R = 0.33$). In the second step, the three model use behaviors explained an additional and statistically significant 12% of the variance in translation accuracy. The partial regression coefficients revealed that mental rotation performance but not abstract reasoning explained a statistically significant portion of the variance ($sr^2 = 0.08$) in the first step. In the second step, both mental rotation performance ($sr^2 = 0.07$) and match target ($sr^2 = 0.10$) uniquely accounted for a statistically significant portion of the variance after controlling for all other variables.

What predicts performance after a delay? A third stepwise multiple regression analysis assessed predictors of diagram translation performance in the delayed posttest. In the first step (see the right portion of Table 9), the individual difference measures explained a statistically significant 11% portion of the variance ($R = .33$). In the second step, the model-use behaviors explained an additional and statistically significant 13% of the variance in translation accuracy. The partial regression coefficients revealed that mental rotation performance ($sr^2 = 0.07$) was a statistically significant predictor of translation accuracy in the first step with both mental rotation performance ($sr^2 = 0.06$) and match target ($sr^2 = 0.08$) each accounting for a statistically significant portion of the variance after controlling for all other variables.

In summary, as in Study 1, interventions with model-based feedback were more successful (86% accuracy in the models posttest) than an intervention with verbal feedback alone (55% accuracy in the models posttest), and this benefit was seen after a week-long delay when models were not available (62% accuracy for the models interventions vs. 48% for the verbal feedback intervention). As in Study 1 and contrary to predictions, there was no difference in posttest performance between those who received concrete and virtual models, indicating that high congruence virtual models and concrete models are equally effective. Finally, as before, using the models to physically enact the desired orientation and conformation of the target diagram predicted performance when models were available and later performance when models were no longer available, consistent with the view that more use of external models leads to more ability to perform the translations internally. Finally, mental rotation performance was more predictive of translation accuracy when models were not available.

General Discussion

This research investigated the development of representational competence in organic chemistry through use of concrete and virtual 3D models to help students learn to translate between multiple 2D diagram formats. The results replicate earlier work (Stull et al., 2013; Padalkar & Hegarty, 2014) in showing that students are more successful in translating between diagrams when they have models available, that using a model to enact the translation process in the world is predictive of learning, and that a model-feedback intervention dramatically improved learning. The new and important contributions of this study are the demonstrations that (1) model-based feedback is superior to verbal-feedback alone, (2) models scaffold learning rather than acting as a crutch, (3) learning with model-based instruction is resilient over a delay of several days, and (4) learning with models transfers to performance when models are no longer available. Finally, our results show that concrete models are equivalent to virtual models in promoting learning, and that the level of action-congruence of the interface to the virtual model does not affect learning.

First, this research demonstrates that model-based feedback is superior to verbal feedback for our task, observed as a large effect size for the models posttest in both studies (Study 1: $r = .67$; Study 2: $r = .57$). Accuracy ranged from 30% to 40% across conditions on the pretest, the model groups were more than 80% accurate on the models posttests, in contrast with the control groups, who were approximately 50% accurate (see Table 3 for exact numbers). While previous research showed a larger effect size for model-based feedback than for verbal feedback in the context of the same

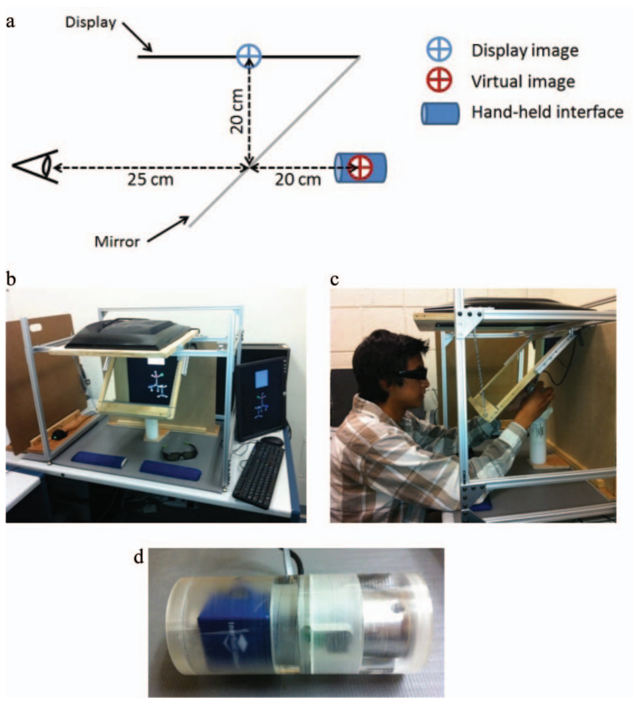


Figure 5. High-congruence virtual reality system. (a) The computer display was mounted horizontally on a metal frame above the desk surface and facing downward. (b) The angled mirror allowed the participant to view the virtual model. (c) The hand-held interface was positioned on an adjustable stand behind the mirror and in the same physical location as the virtual model. (d) The interface enabled the participant to manipulate the virtual model. (Originally published in Stull, et al, 2013 and Barrett, et al, 2014) See the online article for the color version of this figure.

tasks (Padalkar & Hegarty, 2014), that comparison was across experiments. One novel contribution of the present research is to replicate the benefits of model-based feedback over verbal feedback when they are directly compared in the same experiment. Padalkar and Hegarty suggested two possible reasons why model-based feedback intervention was successful, first, that it confronts students' illusion of understanding and second, that it exposes students to the benefits of models. The verbal feedback condition in this experiment confronted students' illusions of understanding and was somewhat successful in improving diagram translation. However, the model-based feedback significantly elevated performance relative to verbal feedback alone, consistent with the idea that model-based feedback relieves cognitive load, enabling students to invest their cognitive resources in learning the representational conventions and translation processes.

Second, models scaffolded learning. Specifically, the superior performance of the models groups when models were no longer available is inconsistent with the models-as-crutch account in that the improved performance in delayed testing for the model conditions indicates a benefit of models for long-term and not just immediate learning. Participants in the models groups in Experiment 1 had 60% to 70% accuracy on the no-models posttest, compared to about 50% for the control group (see Table 3 for exact numbers), observed as a medium effect size for Study 1 ($r = .34$). In contrast, the observed advantage of the models group over the control group was not evident in the no-model posttest in Study 2. However in the delayed posttest of both studies, the models groups performed at 60% to 70% accuracy compared with about 50% for the control group, observed as a medium effect size in both studies (Study 1: $r = .26$; Study 2: $r = .24$). These results demonstrate that models offer more than a vehicle to externalize a process or read-off an answer. We propose that when cognitive load is reduced, the experience of matching the models to the given diagrams promotes students' ability to make referential connections between different representations and that manipulating models and observing the resulting spatial transformations enables students to internalize a mental model of these transformations so that they can later perform these transformations without the scaffold of models. This account is consistent with basic research on mental and manual rotation (Adams et al., 2014; Pani et al., 2005; Smith & Olkun, 2005; Wohlschlager & Wohlschlager, 1998), with principles for the design of chemical visualizations proposed by Wu and Shah (2004) and with theories of enactive learning more generally (Cohen, 1989; Engelkamp et al., 1994; Glenberg et al., 2004; Schwartz & Plass, 2014).

Third, we observed that students who used the external transformation strategy more when models were available were also much better at translation accuracy later when models were not available. Model-use behavior, such as match target, is predictive of successful diagram translation when models are present, because models externalize what would otherwise be a difficult internal manipulation. Critically, this behavior is also predictive of successful diagram translation in delayed tests and when models are not available. This result is further support for the view that performing this task externally leads to the internalization of the relevant cognitive processes.

Fourth, our findings are consistent with previous research (Bodner & McMillan, 1986; Harle & Towns, 2010; Pribyl & Bodner, 1987) in demonstrating a relationship between spatial ability and performance of an organic chemistry task. However, we find that this relationship is attenuated when models are available, and model use is a better predictor of performance than either spatial ability or general intelligence, which was assessed by an abstract reasoning test. This result supports our contention that external models offer a direct representation of 3D space, which is difficult for some students to imagine, especially for students with poor spatial ability, which is often a reflection of limited spatial working memory (Shah & Miyake, 1996).

Fifth, we predicted that action-congruence would affect learning with models, because low-congruence virtual models impose added cognitive load due to a mismatch between the actions performed with the interface and the resulting movements of the models. Although model use took more time with low-congruence virtual models than with concrete models in Study 1, likely reflecting this cognitive load, we did not observe a statistically significant difference between the two model types in learning. Moreover, there was no difference in time or accuracy between concrete models and high-congruence models in Study 2. These results make it clear that action-congruence is not necessary for models to be effective in the present context. This conclusion is consistent with research showing that people can develop mental models that link actions and intended outcomes even when they are not congruent (Schwartz & Holton, 2000), provided that the observed outcome is predictable from the performed action. It is also important to note that in both experiments, learning gains were equivalent with virtual and concrete models, indicating that the additional shape information offered by manipulating concrete models was not necessary for learning.

Table 8

Means, Standard Deviations, and Inferential Statistics Comparing Individual Difference Measures for the Three Intervention Conditions in Study 2

Measures	Condition	*M* (*SD*)	*n*	Inferential test
Age	Concrete	19.72 (.74)	34	
	Virtual	19.59 (1.10)	33	$H(2) = 2.02$, $p = .36$
	Control	19.57 (1.09)	37	
MRT	Concrete	37.39 (15.36)	34	
	Virtual	35.35 (17.56)	33	$H(2) = 1.62$, $p = .44$
	Control	33.73 (18.17)	37	
Abstract reasoning	Concrete	22.83 (7.42)	34	
	Virtual	22.12 (7.81)	33	$H(2) = .11$, $p = .95$
	Control	23.20 (6.92)	37	
GPA	Concrete	3.03 (.32)	34	
	Virtual	3.06 (.36)	33	$H(2) = 1.72$, $p = .42$
	Control	3.11 (.37)	37	
Number of O-chem courses	Concrete	1.89 (.40)	34	
	Virtual	1.88 (.41)	33	$H(2) = .26$, $p = .88$
	Control	1.84 (.44)	37	
Number of intervention cycles	Concrete	4.91 (1.52)	34	
	Virtual	4.03 (1.61)	33	$H(2) = 10.59$, $p < .01$
	Control	5.35 (1.90)	37	

Note. A Kruskal-Wallis Test was performed because assumptions of analyses of variance were not met for some measures. MRT = Mental Rotation Test; GPA = grade point average.

Limitations and Future Directions

There are a number of limitations to these studies, which suggest directions for further research. First, the models posttests in these studies were confounded because participants in the models groups had models available at this posttest, whereas the control groups did not. However, the no-models and delayed posttests did not have this confound, and these are equally or more important outcome measures, as ultimately students need to be able to perform representation translation without models available. In future research, a control condition that had models available at this first posttest would avoid this confounded and allow the assessment of model use when following an intervention without models.

Second, although benefits from the model-based interventions persisted in delayed posttests, it is important to note that performance on the no-models posttests was poorer than on the models posttest, suggesting that gains from the models interventions did not fully transfer to solving problems without models. In a related

Table 9

Hierarchical Multiple Regression Analyses Predicting Diagram Translation Accuracy From Individual Difference Measures (Step 1) and Model-Use Behaviors (Step 2) for Study 2

	Models posttest			No-models posttest			Delayed posttest		
Predictors	*B*	*SE*	β	*B*	*SE*	β	*B*	*SE*	β
Step 1									
Constant	.78	.08		.38	.12		.40	.11	
Mental rotation	<.01	<.01	.19	<.01	<.01	.33*	.01	<.01	.31*
Abstract reasoning	<−.01	<.01	<−.01	<−.01	.01	<−.01	<.01	.01	.04
R^2		.04			.11			.11	
F		1.17			3.80*			3.85*	
Step 2									
Constant	.35	.12		.29	.24		.05	.23	
Mental rotation	<.01	<.01	.18	.01	<.01	.30*	.01	<.01	.29*
Abstract reasoning	<−.01	<.01	−.14	<−.01	.01	−.04	<−.01	.01	−.01
Match start	<.01	.01	.03	.02	.02	.17	.01	.01	.06
Rotate bond	−.01	.02	−.14	−.05	.03	−.30	−.01	.03	−.06
Match target	.06	.01	.77***	.05	.02	.42**	.04	.02	.38*
R^2		.51			.23			.24	
ΔR^2		.48			.12			.13	
$F\Delta R^2$		19.57***			3.20*			3.32*	

Note. $N = 68$. Data conformed to statistical assumptions of the regression analysis.
* $p < .05$. ** $p < .01$. *** $p < .001$.

study, Stieff, Lira, and DeSutter (2014) gave students minimal instructions with models and observed that students performed more poorly in later testing without models. These results, combined with ours, raise questions about the amount of experience with models that is necessary to produce persistent learning outcomes. These questions could be addressed in future studies that controlled the amount of experience with models.

Third, the educational gains in this study were specific to learning to translate between representations. Although acquiring representational competence is a crucial aspect of chemistry education, it is also important to realize that representations are just a tool for reasoning about the characteristics of chemicals, such as stereochemical relations between molecules and how a molecule's shape affects its behavior in chemical reactions (Bucat & Mocerino, 2009). An important next step is to investigate the value of model-based interventions to support students as they reason about molecular structure and reactivity.

A fourth limitation of these studies is that they were conducted in the controlled environment of the psychology laboratory with a one-on-one interaction between an experimenter and a student. Future research should explore whether and how model-based interventions can be effectively implemented in authentic classroom settings. In classroom settings, lecturers typically use models to demonstrate complex concepts without providing students with an opportunity to manipulate models for themselves. This raises the question of whether students can benefit from watching another person manipulating a model, or if they must perform the enactment themselves. Investigating this question, Springer (2014) compared performance of two groups in an organic chemistry course, one receiving a model-based demonstration with verbal explanation of diagrammed concepts and a second with only verbal explanations and diagrams. Students in the model-based demonstration were more successful on posttests than those who received only verbal instructions, suggesting that watching someone else manipulate a model is somewhat effective for learning. However, the present research suggests that active manipulation of the models is likely to provide a benefit beyond just viewing another's actions, which should be investigated further.

Theoretical Implications

Several unique theoretical implications can be derived from this study. First, external 3D models can serve as cognitive scaffolds to support students as they reason about complex spatial problems. Models help to ease cognitive load by externalizing the spatial representations that would otherwise drain cognitive resources for learning. Second, enactive use of 3D models can also support students in learning to translate between representations and reason about spatial processes. Most of the prior work on enactive learning has focused verbal learning such as recall and recognition of phrases (Cohen, 1989) or reading comprehension (Glenberg et al., 2004). The present study demonstrates that enactment can also offer value in educating complex spatial skills, in a STEM domain. More generally, augmenting cognition with external representations can enable students to construct spatial mental models for later use in problem solving. Furthermore, reasoning about spatial processes by manipulating external models is not dependent on the congruence between the performed interface actions and the observed motion of the models. Factors such as action-congruence and haptic fidelity of the hand-held interface, have no effect on learning in this context. This suggests that the value of model enactment is due to intentional interactions with models, or visually observing the results of these interactions.

Educational Implications

This study has several educational implications. First, a clear implication is that it is important to incorporate manipulation of models into instruction in organic chemistry. Manipulating models does not just offload difficult mental transformation processes on external actions, but can lead to the internalization of these transformations and therefore train mental transformation processes. Furthermore, using models as feedback is a particularly effective way of inducing students to engage with models and experience their benefits, and this has been shown not just in organic chemistry, but also in geometry (Cohen & Hegarty, 2014) and geology (Gagnier, Atit, Ormand, & Shipley, 2012) indicating that it may be a general strategy that can be used for teaching in spatially rich domains. Finally, our research suggests that it does not matter whether models are virtual or concrete or employ a high- or low-congruence interface as long as students interact with and receive feedback from the models they use as tools for enactment, providing new evidence for the potential of virtual resources in education.

References

Abrahamson, D., & Lindgren, R. (2014). Embodiment and embodied design. In R. K. Sawyer (Ed.), *Cambridge handbook of the learning sciences* (2nd ed., pp. 358–376). New York, NY: Cambridge University Press.

Adams, D. M., Stull, A. T., & Hegarty, M. (2014). Effects of mental and manual rotation training on mental and manual rotation performance. *Spatial Cognition and Computation, 14,* 169–198. http://dx.doi.org/10.1080/13875868.2014.913050

Ainsworth, S. E. (1999). The functions of multiple representations. *Computers & Education, 33,* 131–152. http://dx.doi.org/10.1016/S0360-1315(99)00029-9

Barak, M., & Dori, Y. J. (2004). Enhancing undergraduate students' chemistry understanding through project-based learning in an IT environment. *Science Education, 89,* 117–139. http://dx.doi.org/10.1002/sce.20027

Barnett, M., Yamagata-Lynch, L., Keating, T., Barab, S. A., & Hay, K. E. (2005). Using virtual reality computer models to support student understanding of astronomical concepts. *Journal of Computers in Mathematics and Science Teaching, 24,* 333–356.

Barrett, T. J., Stull, A. T., Hsu, T. M., & Hegarty, M. (2014). Constrained interactivity for relating multiple representations in science: When virtual is better than real. *Computers & Education, 81,* 69–81. http://dx.doi.org/10.1016/j.compedu.2014.09.009

Bennett, G. K., Seashore, H. G., & Wesman, A. G. (1974). *Manual for the Differential Aptitude Test* (5th ed.). New York, NY: Psychological Corporation.

Bodner, G. M., & McMillan, T. L. B. (1986). Cognitive restructuring as an early stage in problems solving. *Journal of Research in Science Teaching, 23,* 727–737. http://dx.doi.org/10.1002/tea.3660230807

Bruner, J. S. (1957). *Going beyond the information given*. New York, NY: Norton.

Bucat, B., & Mocerino, M. (2009). Learning at the sub-micro level: Structural representations. In J. K. Gilbert & D. Treagust (Eds.), *Multiple representations in chemical education: Models and modeling in science education* (Vol. 4, pp. 11–29). Dordrecht, NL: Springer. http://dx.doi.org/10.1007/978-1-4020-8872-8_2

Cary, M., & Carlson, R. A. (2001). Distributing working memory resources during problem solving. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 27,* 836–848. http://dx.doi.org/10.1037/0278-7393.27.3.836

Chariker, J. H., Naaz, F., & Pani, J. R. (2011). Computer-based learning of neuroanatomy: A longitudinal study of learning, transfer, and retention. *Journal of Educational Psychology, 103,* 19–31. http://dx.doi.org/10.1037/a0021680

Cheng, M., & Gilbert, J. K. (2009). Towards a better utilization of diagrams in research into the use of representative levels of chemical education. In J. K. Gilbert & D. Treagust (Eds.), *Multiple representations in chemical education: Models and modeling in science education. Dordrecht,* NL: Springer. http://dx.doi.org/10.1007/978-1-4020-8872-8_4

Cohen, C. A., & Hegarty, M. (2014). Visualizing cross-sections: Training spatial thinking using interactive animations and virtual objects. *Learning and Individual Differences, 33,* 63–71. http://dx.doi.org/10.1016/j.lindif.2014.04.002

Cohen, R. L. (1989). Memory for action events: The power of enactment. *Educational Psychology Review, 1,* 57–80. http://dx.doi.org/10.1007/BF01326550

Copolo, C. E., & Hounshell, P. B. (1995). Using three-dimensional models to teach molecular structures in high school chemistry. *Journal of Science Education and Technology, 4,* 295–305. http://dx.doi.org/10.1007/BF02211261

Dori, Y. J., & Kaberman, K. (2012). Assessing high school chemistry students' modeling sub-skills in a computerized molecular modeling learning environment. *Instructional Science: An International Journal of the Learning Sciences, 40,* 69–91. http://dx.doi.org/10.1007/s11251-011-9172-7

Engelkamp, J., Zimmer, H. D., Mohr, G., & Sellen, O. (1994). Memory of self-performed tasks: Self-performing during recognition. *Memory & Cognition, 22,* 34–39. http://dx.doi.org/10.3758/BF03202759

Ernst, M. O., & Banks, M. S. (2002). Humans integrate visual and haptic information in a statistically optimal fashion. *Nature, 415,* 429–433. http://dx.doi.org/10.1038/415429a

Ferk, V., Vrtacnik, M., Blejec, A., & Gril, A. (2003). Students' understanding of molecular structure representations. *International Journal of Science Education, 25,* 1227–1245. http://dx.doi.org/10.1080/09500690220000038231

Field, A. (2005). *Discovering statistics using SPSS*. London, England: Sage.

Francoeur, E., & Segal, J. (2004). From model kits to interactive computer graphics. In S. de Chadarevian & N. Hopwood (Eds.), *Models: Third dimensions of science* (pp. 402–429). Stanford, CA: Stanford University Press.

Gagnier, K. M., Atit, K., Ormand, C., & Shipley, T. F. (2012, November). *Improving penetrative thinking via progressive alignment and directed sketching*. Paper presented at The Geological Society of America Annual Conference, Charlotte, NC.

Gentner, D. (1983). Structure-mapping: A theoretical framework for analogy. *Cognitive Science, 7,* 155–170. http://dx.doi.org/10.1207/s15516709cog0702_3

Glenberg, A. M., Gutierrez, T., Levin, J. R., Japuntich, S., & Kaschak, M. P. (2004). Activity and imagined activity can enhance young children's reading comprehension. *Journal of Educational Psychology, 96,* 424–436. http://dx.doi.org/10.1037/0022-0663.96.3.424

Goodwin, W. M. (2008). Structural formulas and explanation in organic chemistry. *Foundations of Chemistry, 10,* 117–127. http://dx.doi.org/10.1007/s10698-007-9033-2

Harle, M., & Towns, M. (2010). A review of spatial ability literature, its connection to chemistry, and implications for instruction. *Journal of Chemical Education, 88,* 351–360. http://dx.doi.org/10.1021/ed900003n

Harris, M. A., Peck, R. F., Colton, S., Morris, J., Chaibub Neto, E., & Kallio, J. (2009). A combination of hand-held models and computer imaging programs helps students answer oral questions about molecular structure and function: A controlled investigation of student learning. *CBE Life Sciences Education, 8,* 29–43. http://dx.doi.org/10.1187/cbe.08-07-0039

Kali, Y., & Orion, N. (1996). Spatial abilities of high-school students in the perception of geologic structures. *Journal of Research in Science Teaching, 33,* 369–391. http://dx.doi.org/10.1002/(SICI)1098-2736(199604)33:4<369::AID-TEA2>3.0.CO;2-Q

Kalyuga, S. (2007). Enhancing instructional efficiency of interactive e-learning environments: A cognitive load perspective. *Educational Psychology Review, 19,* 387–399. http://dx.doi.org/10.1007/s10648-007-9051-6

Kaminski, J. A., Sloutsky, V. M., & Heckler, A. (2009). Transfer of mathematical knowledge: The portability of generic instantiations. *Child Development Perspectives, 3,* 151–155. http://dx.doi.org/10.1111/j.1750-8606.2009.00096.x

Kastens, K. A., & Rivet, A. (2010). Using analogical mapping to assess the affordances of scale models. In C. Holscher, T. F. Shipley, M. O. Belardinelli, J. A. Bateman, & N. S. Newcombe (Eds.), *Spatial cognition, VII* (pp. 112–124). Heidelberg, Germany: Springer. http://dx.doi.org/10.1007/978-3-642-14749-4

Keig, P. F., & Rubba, P. (1993). Translation of representations of the structure of matter and its relationship to reasoning, gender, spatial reasoning, and specific prior knowledge. *Journal of Research in Science Teaching, 30,* 883–903. http://doi.wiley.com/10.1002/tea.3660300807. http://dx.doi.org/10.1002/tea.3660300807

Kirsh, D. (1995). The intelligent use of space. *Artificial Intelligence, 73,* 31–68. http://dx.doi.org/10.1016/0004-3702(94)00017-U

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

Novick, L. R., Stull, A. T., & Catley, K. M. (2012). Reading phylogenetic trees: The effects of tree orientation and text processing on comprehension. *Bioscience, 62,* 757–764. http://dx.doi.org/10.1525/bio.2012.62.8.8

Padalkar, S., & Hegarty, M. (2014). Models as feedback: Developing representational competence in chemistry. [Advance online publication.] *Journal of Educational Psychology*.

Pani, J. R., Chariker, J. H., Dawson, T. E., & Johnson, N. (2005). Acquiring new spatial intuitions: Learning to reason about rotations. *Cognitive Psychology, 51,* 285–333. http://dx.doi.org/10.1016/j.cogpsych.2005.06.002

Pauwels, L. (2006). Introduction: The role of visual representations in the production of scientific reality. In L. Pauwels (Ed.), *Visual cultures of science: Rethinking representational practices in knowledge building and science communications* (pp. vii–xix). Lebanon, NH: Dartmouth College Press.

Preece, D., Williams, S. B., Lam, R., & Weller, R. (2013). "Let's get physical": Advantages of a physical model over 3D computer models and textbooks in learning imaging anatomy. *Anatomical Sciences Education, 6,* 216–224. http://dx.doi.org/10.1002/ase.1345

Pribyl, J. R., & Bodner, G. M. (1987). Spatial ability and its role in organic chemistry: A study of four organic courses. *Journal of Research in Science Teaching, 24,* 229–240. http://dx.doi.org/10.1002/tea.3660240304

Rosnow, R. L., Rosenthal, R., & Rubin, D. B. (2000). Contrasts and correlations in effect-size estimation. *Psychological Science, 11,* 446–453. http://dx.doi.org/10.1111/1467-9280.00287

Rozenblit, L., & Keil, F. (2002). The misunderstood limits of folk science: An illusion of explanatory depth. *Cognitive Science, 26,* 521–562. http://dx.doi.org/10.1207/s15516709cog2605_1

Ruddle, R., & Jones, D. M. (2001). Manual and virtual rotation of a three-dimensional object. *Journal of Experimental Psychology*. Applied, *7,* 286–96. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/11838891.

Satava, R. M. (1999). Emerging technologies for surgery in the 21st century. *Archives of Surgery, 134,* 1197–1202. http://dx.doi.org/10.1001/archsurg.134.11.1197

Savec, V. F., Vrtacnik, M., & Gilbert, J. K. (2005). *Evaluating the educational value of molecular structure representations*. Dordrecht, NL: Springer. http://dx.doi.org/10.1007/1-4020-3613-2_14

Schwartz, D. L., & Holton, D. L. (2000). Tool use and the effect of action on the imagination. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26,* 1655–1665. http://dx.doi.org/10.1037/0278-7393.26.6.1655

Schwartz, R. N., & Plass, J. L. (2014). Click versus drag: User-performed tasks and the enactment effect in an interactive multimedia environment. *Computers in Human Behavior, 33,* 242–255. http://dx.doi.org/10.1016/j.chb.2014.01.012

Shah, P., & Miyake, a. (1996). The separability of working memory resources for spatial thinking and language processing: An individual differences approach. *Journal of Experimental Psychology*. General, *125,* 4–27. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/8851737.

Smith, G. G., & Olkun, S. (2005). Why interactivity works: Interactive priming of mental rotation. *Journal of Educational Computing Research, 32,* 93–111. http://dx.doi.org/10.2190/4KA5-03UX-A70E-E53W

Snyder, C. W., Vandromme, M. J., Tyra, S. L., & Hawn, M. T. (2009). Proficiency-based laparoscopic and endoscopic training with virtual reality simulators: A comparison of proctored and independent approaches. *Journal of Surgical Education, 66,* 201–207. http://dx.doi.org/10.1016/j.jsurg.2009.07.007

Springer, M. T. (2014). Improving students' understanding of molecular structure through broad-based use of computer models in the undergraduate organic chemistry lecture. *Journal of Chemical Education, 91,* 1162–1168. http://dx.doi.org/10.1021/ed400054a

Stieff, M., Lira, M., & DeSutter, D. (2014). Representational competence & spatial thinking in STEM. *Proceedings of the 12th International Conference of the Learning Sciences*. Boulder, CO: ISLS.

Stull, A. T., Barrett, T. J., & Hegarty, M. (2013). Usability of concrete and virtual models in chemistry instruction. *Computers in Human Behavior, 29,* 2546–22556. http://dx.doi.org/10.1016/j.chb.2013.06.012

Stull, A. T., Hegarty, M., Dixon, B., & Stieff, M. (2012). Representational translations with concrete models in organic chemistry. *Cognition and Instruction, 30,* 404–434. http://dx.doi.org/10.1080/07370008.2012.719956

Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science, 12,* 257–285. http://dx.doi.org/10.1207/s15516709cog1202_4

Triona, L. M., & Klahr, D. (2003). Point and click or grab and heft: Comparing the influence of physical and virtual instructional materials on elementary school students' ability to design experiments. *Cognition and Instruction, 21,* 149–173. http://dx.doi.org/10.1207/S1532690XCI2102_02

Trumbo, J. (2006). Making science visible: Visual literacy in science communication. In L. Pauwels (Ed.), *Visual cultures of science: Rethinking representational practices in knowledge building and science communications* (pp. 266–309). Lebanon, NH: Dartmouth College Press.

Tversky, B. (2005). Visuospatial reasoning. In K. J. Holyoak & R. Morrison (Eds.), *The Cambridge handbook of thinking and reasoning* (pp. 209–241). New York, NY: Cambridge University Press.

U.S. Government Accountability Office. (2005). *College textbooks: Enhancing offerings appear to drive recent price increases* (GAO-05–806). Retrieved from http://www.gao.gov/assets/250/247332.pdf

Vandenberg, S., & Kuse, A. (1978). Mental rotation, a group test of 3-D spatial visualization. *Perceptual and Motor Skills, 47,* 599–604. http://dx.doi.org/10.2466/pms.1978.47.2.599

Wexler, M., Kosslyn, S. M., & Berthoz, A. (1998). Motor processes in mental rotation. *Cognition, 68,* 77–94. http://dx.doi.org/10.1016/S0010-0277(98)00032-8

Wiedenbauer, G., & Jansen-Osmann, P. (2008). Manual training of mental rotation in children. *Learning and Instruction, 18,* 30–41. http://dx.doi.org/10.1016/j.learninstruc.2006.09.009

Wohlschläger, A., & Wohlschläger, A. (1998). Mental and manual rotation. *Journal of Experimental Psychology: Human Perception and Performance, 24,* 397–412. http://dx.doi.org/10.1037/0096-1523.24.2.397

Wu, H., Krajcik, J. S., & Soloway, E. (2001). Promoting understanding of chemical representations: Students' use of a visualization tool in the classroom. *Journal of Research in Science Teaching, 38,* 821–842. http://dx.doi.org/10.1002/tea.1033

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

Yelland, N., & Masters, J. (2007). Rethinking scaffolding in the information age. *Computers & Education, 48,* 362–382. http://dx.doi.org/10.1016/j.compedu.2005.01.010

(*Appendices follow*)

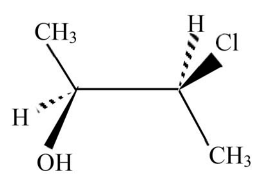
Appendix A

Instructions I

Welcome and thank you for agreeing to participate in our study! It is about learning with molecular representations in organic chemistry. In this study, you will complete several worksheets. For each, you will be shown a Newman or a Dash-Wedge diagram of a molecule. Your task will be to draw a different diagram for each molecule. For example, you might be given a Dash-Wedge diagram of a molecule and asked to draw the corresponding Newman diagram for the same molecule. The text on each page will describe which diagram you are to draw. Some of the transformations may be difficult but please try your best.

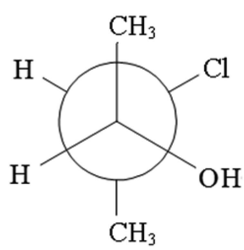
Before we proceed to the worksheets, we will review the rules for interpreting the different diagrams that you will be expected to draw. Both use different conventions to illustrate the 3D shape of the molecule. Notice that the same 4-carbon molecule is illustrated in both diagrams in the examples below.

Dash-Wedge



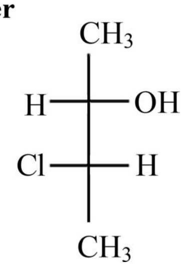
In a Dash-Wedge diagram, the molecule is oriented with the backbone carbons at the two 4-way intersections of lines on the left and right of the diagram. Dashed lines represent bonds to atoms that are going into the page (below the plane of the paper). Wedge lines represent atoms that are coming out of the page (above the plane of the paper). Solid lines represent bonds to atoms that are in the plane of the paper.

Newman



In a Newman diagram, the molecule is oriented with one backbone carbon in front of the other. The front carbon is located at the intersection of the three lines (noon, 4 o'clock, and 8 o'clock around the circle). The atoms at the ends of these three lines are attached to the front carbon. The rear carbon is behind the circle. The atoms at the ends of the shorter lines connected to the circle (2 o'clock, 6 o'clock, and 10 o'clock around the circle) are attached to the rear carbon.

Fischer



In the Fischer diagram the components at the right and left of the horizontal lines are coming out of the page (above the plane of the paper) and the atoms at the top and bottom of the vertical line are going into the page (below the plane of the paper). The two backbone carbons are located where the horizontal lines cross the vertical line. These carbons are on the plane of the paper.

Take a moment to visualize how each diagram represents the three-dimensional structure of the molecule and satisfy yourself that both diagrams represent the same molecule. Compare and contrast the diagrams because you will need to draw each in the following activity.

Please let the experimenter know if you have any questions.

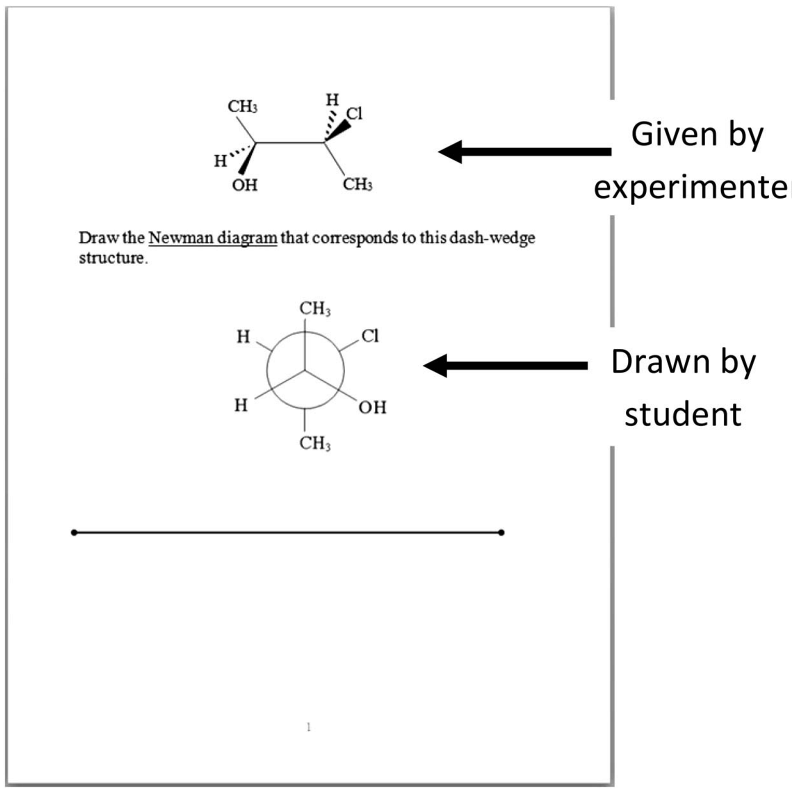
(*Appendices continue*)

Appendix B

Instructions II

Next, you will be given six diagrams. For each of these diagrams, please draw a different diagram of the same molecule above the line that divides the page.

On each worksheet, one diagram will be given at the top of each page. Below each diagram will be instructions indicating which diagram you should draw. Here is an example of a completed worksheet.



The task is not timed so please strive to be accurate.

Please note that you do not have to draw the most energy efficient conformer of the molecule. Also, some of the molecules you will see might be uncommon or unfamiliar to you.

As a reminder:
Carbon is C
Hydrogen is H
Nitrogen is N
Oxygen is O
Chlorine is Cl

Please let the experimenter know if you have any questions.

Received December 12, 2014
Revision received July 20, 2015
Accepted July 21, 2015 ■