

We have presented the Graphical Abstract text and image for your article below. This brief summary of your work will appear in the contents pages of the issue in which your article appears.

---

## **Building mental models of a reaction mechanism: the influence of static and animated representations, prior knowledge, and spatial ability**

Q1

Amanda Bongers, Berthorie Beauvoir, Nicholas Streja, Georg Northoff and Alison B. Flynn\*

---

Please check this proof carefully. Our staff will not read it in detail after you have returned it.

Please send your corrections either as a copy of the proof PDF with electronic notes attached or as a list of corrections. **Do not edit the text within the PDF or send a revised manuscript** as we will not be able to apply your corrections. Corrections at this stage should be minor and not involve extensive changes.

**Proof corrections must be returned as a single set of corrections, approved by all co-authors. No further corrections can be made after you have submitted your proof corrections as we will publish your article online as soon as possible after they are received.**

Please ensure that:

- The spelling and format of all author names and affiliations are checked carefully. You can check how we have identified the authors' first and last names in the researcher information table on the next page. **Names will be indexed and cited as shown on the proof, so these must be correct.**
- Any funding bodies have been acknowledged appropriately and included both in the paper and in the funder information table on the next page.
- All of the editor's queries are answered.
- Any necessary attachments, such as updated images or ESI files, are provided.

Translation errors can occur during conversion to typesetting systems so you need to read the whole proof. In particular please check tables, equations, numerical data, figures and graphics, and references carefully.

Please return your **final** corrections, where possible within **48 hours** of receipt, by e-mail to: [cerp@rsc.org](mailto:cerp@rsc.org). If you require more time, please notify us by email.

## Funding information

Providing accurate funding information will enable us to help you comply with your funders' reporting mandates. Clear acknowledgement of funder support is an important consideration in funding evaluation and can increase your chances of securing funding in the future.

We work closely with Crossref to make your research discoverable through the Funding Data search tool (<http://search.crossref.org/funding>). Funding Data provides a reliable way to track the impact of the work that funders support. Accurate funder information will also help us (i) identify articles that are mandated to be deposited in **PubMed Central (PMC)** and deposit these on your behalf, and (ii) identify articles funded as part of the **CHORUS** initiative and display the Accepted Manuscript on our web site after an embargo period of 12 months.

Further information can be found on our webpage (<http://rsc.li/funding-info>).

### What we do with funding information

We have combined the information you gave us on submission with the information in your acknowledgements. This will help ensure the funding information is as complete as possible and matches funders listed in the Crossref Funder Registry.

If a funding organisation you included in your acknowledgements or on submission of your article is not currently listed in the registry it will not appear in the table on this page. We can only deposit data if funders are already listed in the Crossref Funder Registry, but we will pass all funding information on to Crossref so that additional funders can be included in future.

### Please check your funding information

The table below contains the information we will share with Crossref so that your article can be found *via* the Funding Data search tool. **Please check that the funder names and grant numbers in the table are correct and indicate if any changes are necessary to the Acknowledgements text.**

Funder name	Funder's main country of origin	Funder ID (for RSC use only)	Award/grant number

## Researcher information

Please check that the researcher information in the table below is correct, including the spelling and formatting of all author names, and that the authors' first, middle and last names have been correctly identified. **Names will be indexed and cited as shown on the proof, so these must be correct.**

If any authors have ORCID or ResearcherID details that are not listed below, please provide these with your proof corrections. Please ensure that the ORCID and ResearcherID details listed below have been assigned to the correct author. Authors should have their own unique ORCID iD and should not use another researcher's, as errors will delay publication.

Please also update your account on our online [manuscript submission system](#) to add your ORCID details, which will then be automatically included in all future submissions. See [here](#) for step-by-step instructions and more information on author identifiers.

First (given) and middle name(s)	Last (family) name(s)	ResearcherID	ORCID iD
Amanda	Bongers		
Berthorie	Beauvoir		
Nicholas	Streja		
Georg	Northoff		
Alison B.	Flynn	H-5595-2013	0000-0002-9240-1287

## Queries for the attention of the authors

Journal: **Chemistry Education Research and Practice**

Paper: **c9rp00198k**

Title: **Building mental models of a reaction mechanism: the influence of static and animated representations, prior knowledge, and spatial ability**

For your information: You can cite this article before you receive notification of the page numbers by using the following format: (authors), Chem. Educ. Res. Pract., (year), DOI: 10.1039/c9rp00198k.

Editor's queries are marked on your proof like this **Q1**, **Q2**, etc. and for your convenience line numbers are indicated like this 5, 10, 15, ...

Please ensure that all queries are answered when returning your proof corrections so that publication of your article is not delayed.

Query reference	Query	Remarks
Q1	Please confirm that the spelling and format of all author names is correct. Names will be indexed and cited as shown on the proof, so these must be correct. No late corrections can be made.	
Q2	Ref. Jones (2005, 2013, 2015), Johnson-Laird (1983), Bodner and Domin (2000), Rapp (2005), Gilbert <i>et al.</i> (1998), Clement (2000), Greca and Moreira (2000), Coll (2006), Schwarz <i>et al.</i> (2009), Stieff <i>et al.</i> (2016), Lowe (2008), Scalco <i>et al.</i> (2017), Flynn and Ogilvie (2015), Ogilvie <i>et al.</i> (2017), Flynn (2015), Bodner and Guay (1997) and Yoon (2011) are cited within the text but do not appear to be included in the reference list. Do you wish to add these references to the reference list or would you like the citations to be removed from the text?	
Q3	A citation to Table 2 has been added here, please check that the placement of this citation is suitable. If the location is not suitable, please indicate where in the text the citation should be inserted.	
Q4	Ref. Bongers <i>et al.</i> (2019): Can this reference be updated? If so, please provide the relevant information such as year, volume and page or article numbers as appropriate.	
Q5	Please indicate where ref. Scalco <i>et al.</i> (2018) should be cited in the text.	

10  
Building mental models of a reaction mechanism:  
the influence of static and animated  
representations, prior knowledge, and  
spatial ability Q1 10

Cite this: DOI: 10.1039/c9rp00198k

15  
Amanda Bongers,<sup>a</sup> Berthorie Beauvoir,<sup>a</sup> Nicholas Streja,<sup>a</sup> Georg Northoff<sup>b</sup> and  
Alison B. Flynn \*<sup>a</sup> 15

In chemistry, novices and experts use mental models to simulate and reason about sub-microscopic processes. Animations are thus important tools for learning in chemistry to convey reaction dynamics and molecular motion. While there are many animations available and studies showing the benefit of learning from animations, there are also limitations to their design and effectiveness. Moreover, there are few experimental studies into learning chemistry from animations, especially organic reaction mechanisms. We conducted a mixed-methods study into how students learn and develop mental models of a reaction mechanism from animations. The study ( $N = 45$ ) used a pre-/post-test experimental design and counterbalanced static and animated computerized learning activities (15 min each), plus short think-aloud interviews for some participants ( $n = 20$ ). We developed the tests and learning activities in a pilot study; these contained versions of an epoxide opening reaction mechanism either as static (using the electron-pushing formalism) or animated representations. Participants' test accuracy, response times, and self-reported confidence were analyzed quantitatively ( $\alpha = 0.05$ ) and we found that, while participants showed a learning effect, there were no significant differences between the static and animated learning conditions. Participants' spatial abilities were correlated to their test accuracy and influenced their learning gains for both conditions. Qualitative framework analysis of think-aloud interviews revealed changes in participants' reasoning about the test questions, moving toward using rule- and case-based reasoning over model-based reasoning. This analysis also revealed that dynamic and transitional features were incorporated into participants' working mental models of the reaction mechanism after learning from animations. The divergence of participants' mental models for reasoning and visualization could suggest a gap in their mental model consolidation.

20  
25  
30  
Received 4th September 2019,  
Accepted 9th December 201935  
DOI: 10.1039/c9rp00198k

rsc.li/cerp

40  
Introduction

Visualizations such as animations are often used as teaching models in chemistry designed to go beyond on-paper representations with the aim to help students develop robust mental models of dynamic processes and improve their representational competence (Wu and Shah, 2004; Stieff, 2011a; Stieff *et al.*, 2011; Suits and Sanger, 2013; Suits, 2015) Some studies report that molecular animations help students imagine and understand the sub-microscopic domain and make connections to macroscopic events and changes in matter (Williamson and Abraham, 1995; Stieff and Wilensky, 2003;

Velázquez-Marcano *et al.*, 2004; Ardac and Akaygun, 2005; Tasker and Dalton, 2006; Kelly and Jones, 2007, 2008; Al-Balushi and Al-Hajri, 2014; Ryoo and Linn, 2014; Akaygun, 2016). However, there are also reported challenges with designing and learning from animations (Lowe, 2004; Ardac and Akaygun, 2005; Kelly and Jones, 2008; Tasker and Dalton, 2008; Suits, 2015) and limited controlled studies directly comparing learning from animations compared to traditional static representations. In a meta-review, Höffler (2010) stated that many studies found no general learning gains from teaching students using dynamic animations compared to using static pictures.

The design and features of animations for learning must be chosen with care, since learners may focus on salient rather than relevant features which can result in the development or reinforcement of misconceptions (Lowe, 2004; Tasker and Dalton, 2008; Kelly *et al.*, 2017). Suits (2013, 2015) and Jones

45  
55  
<sup>a</sup> Department of Chemistry and Biomolecular Sciences, University of Ottawa, Ottawa, Ontario, Canada. E-mail: Alison.Flynn@uOttawa.ca<sup>b</sup> Institute of Mental Health Research, University of Ottawa, Ottawa, Ontario, Canada

1 (2005, 2013, 2015) have led key research and development into  
designing animations for learning chemistry concepts, and  
caution that animations can contain extraneous information  
are often more complex in features than the comparative static  
5 material. For example, chemistry animations typically use dif-  
ferent types of symbols or visuals than traditional models to  
represent atoms, bonds, or processes. Novices can struggle to  
move between these multiple representations though experts do  
so easily (Kozma and Russell, 1997; Kozma, 2003; Taber, 2013).

10 Many studies into learning from chemistry animations use a  
qualitative lens and there is a need for more research using  
mixed methods to allow comparisons between qualitative find-  
ings and quantitative results. Currently there are limited stud-  
ies into chemistry animations that use controlled  
15 experimental design. One experimental study by Kelly and  
Jones (2008) compared first year general chemistry students'  
ability to transfer of knowledge from VisChem animations and  
animations from their textbook. The authors reported that the  
students learned to incorporate features from the animations  
20 about dissolved salt into their mental models and drawings and  
found a wide variation in their mental models and prior  
knowledge used to construct their models. An advantage with  
their experimental design was that each student participated in  
both treatments, allowing the authors to compare learning in  
25 individuals as well as across treatments. More recently, Akay-  
gun (2016) reported on significant changes in 10–11th grade  
students' mental models of the oxygen atom after they gener-  
ated an animation with software. This study used a pre- and  
post-test design where students drew static representations of  
30 the oxygen atom; the authors found that the modelling activity  
increased the number of dynamic features in the static models.

Of these studies on learning from chemistry animations,  
only a few report the effects of learning organic reaction  
mechanisms from animations, despite widespread availability  
and use in courses or online (ChemTube3D: Organic  
35 Chemistry Animations, 2018; Organic ChemWare, 2018; Khan  
Academy: Organic chemistry, 2019). Organic reaction mecha-  
nisms are complex dynamic processes, where aspects of mole-  
cular and particulate motion are not conveyed in traditional  
representations. For example, the transfer of electrons in the  
40 reaction is portrayed statically using arrows of the electron-  
pushing formalism (EPF) (Bhattacharyya, 2013). Our prior work  
showed that students had dynamic mental models of reactivity  
as particles in motion and suggested that prior viewing of  
chemistry animations may have cued students to use dynamic  
45 mental models in later problem-solving situations (Bongers  
*et al.*, 2019a, 2019b). Aldahmash and Abraham (2009) showed  
that students performed better on a test about organic reaction  
mechanisms after viewing three-dimensional animations on a  
computer compared to those who only viewed two-dimensional  
50 static images, although the design and features of these dif-  
ferent representations are unclear and not discussed. In their  
study, prior knowledge was measured with a generalized test of  
the course content knowledge and was used to ensure no  
55 significant differences between the two participant groups.  
Another study using a post-test only design found that students

who used animations of molecules and reactions along with  
concrete models performed significantly better on a post-test  
1 compared to a group who only used concrete models, and prior  
knowledge was assumed to be equal in both groups  
(Al-Balushi and Al-Hajri, 2014). However, in order to experi-  
5 mentally measure learning differences between static or ani-  
mated conditions in individuals, a controlled pre-test and post-  
test design is required. Baptista *et al.* (2019) used pre/post-test  
design to study a group of students learning the saponification  
reaction from multiple representations. They used a word-  
10 association task as their test and found that students' mental  
models were influenced by the sequence of lessons containing  
videos, a laboratory activity, symbolic and sub-microscopic  
explanations, and concrete models. This controlled pre/post  
15 design is especially important in studies of mental models,  
where an individual's prior knowledge is the foundation for  
mental model building (Lowe and Boucheix, 2008). Herein we  
used a mixed methods experimental design to explore and  
compare how individual students learn and develop mental  
20 models of a reaction mechanism from traditional static repre-  
sentations and complementary animations.

Learning from animations may also be influenced by the  
individual learner's spatial ability. According to Aldahmash and  
Abraham (2009), students with high spatial ability who were  
25 taught using their three-dimensional animations displayed  
greater learning gains when compared to individuals with high  
spatial ability who were taught using two-dimensional static  
images. In contrast, Höffler and Leutner (2011) found that  
learners with low spatial ability learned better from animations  
while individuals with high spatial ability learned better from  
30 static images. These contrasting findings indicate that the type  
and design of the animation, and the assessment used to study  
learning, will influence the role of spatial ability. Spatial ability  
is an essential skill used frequently in organic chemistry, for ex-  
ample when transitioning between two- to three-dimensional repre-  
35 sentations of molecules (Harle and Towns, 2011) or when  
questions require students to mentally manipulate molecules  
(Pribyl and Bodner, 1987). This being said, it is still unknown as  
to whether or not an individual's spatial ability is an active  
predictor of their success in organic chemistry (Stieff *et al.*, 2012).  
40

### Research goals and questions

We conducted a mixed methods study into how students learn  
from traditional static representations of a reaction mechanism  
45 compared to complementary (simplified) Organic Chemware<sup>®</sup>  
animations, which are used in chemistry courses at our institu-  
tion. This study stemmed from our prior work on students'  
working mental models of organic reaction mechanisms and  
the EPF. Both the static representations using the EPF and  
50 animations are two different types of models of reaction  
mechanisms; our aim was to study how learning from different  
models influences an individual's mental models. Our guiding  
research questions were:

1. How does learning a reaction mechanism from static or  
55 animated visualizations affect measures of learners' test accu-  
racy, response times, and confidence?

1 2. What is the influence of the learner's spatial ability on the  
above measures?

3 3. How does learning a reaction mechanism from static or  
5 animated visualizations affect learners' working mental models  
of the reaction?

## Theoretical framework

10 We used methods and frameworks derived from the reasoning  
literature and mental models theory to investigate how learners  
develop and use their mental models of a reaction mechanism  
(Johnson-Laird, 1983). Mental models are an individual's personal  
15 knowledge structures for representing how things work in  
their external reality (Bodner and Domin, 2000; Rapp, 2005).  
Mental models theory has been important in science education  
and visualization literature, due to field's reliance on models  
and modelling (Gilbert *et al.*, 1998; Clement, 2000; Greca and  
20 Moreira, 2000; Coll, 2006; Schwarz *et al.*, 2009; Stieff, 2011a,  
2011b; Stieff *et al.*, 2016).

### Learning from animated diagrams

25 Lowe and Boucheix's (2008) framework outlined five phases for  
how learners process complex animations to develop their  
mental models: (1) localized perceptual exploration, (2) regional  
structure formation, (3) global characterization, (4) functional  
differentiation, and (5) mental model consolidation  
(Appendix, Table 5). This framework outlines information  
30 processing from animations as both "bottom-up" perception  
and "top-down" incorporation of prior general and domain-  
specific knowledge. The early phases involve perception and  
finding causal spatiotemporal links within the animation,  
while later stages describe recruitment and consolidation of  
35 prior knowledge with the developing mental model. Lowe  
(2008) explains how the later phases in his framework are the  
most challenging for domain novices, and stresses the need for  
more research on the learner's prior knowledge in building  
mental models.

40 Lowe and Boucheix's framework applies generally to ani-  
mated diagrams, which can range from fluid (high frame rate)  
videos to a series of pictures in shown quick succession  
(Lowe, 1999; Scalco *et al.*, 2017). Due to the design of experi-  
ment and the learning materials in the study herein (see  
45 Appendix, Fig. 11), the phases are relevant to both the static  
and animated learning conditions, and we applied this frame-  
work to our analysis of the effect of animations and to the  
overall discussion of the results.

### Working mental models and reasoning

50 In the present study, we used a coding scheme developed in a  
previous qualitative study for describing organic chemistry  
students' mental models of the epoxide-opening reaction  
mechanism (Bongers *et al.*, 2019a, 2019b). In that prior work  
we found that the participants had and used different types of  
55 working mental models in their descriptions of the reaction  
mechanisms they envision and provided on paper. These

working mental models (Appendix, Table 5) were static, 1  
dynamic (process), or dynamic (particles in motion). Partici-  
pants were found to use more than one type of mental model  
during the interview depending on the context (*e.g.*, worksheet  
5 question or interview prompt). In particular, we found that  
participants used dynamic working mental models when  
prompted to describe how they visualized the reaction.

Mental models are used for visualization, simulation, and also  
reasoning about systems. In this study, we were interested in not  
10 only characterizing the nature of students' working mental  
models, but how (or if) students used their visualization models  
for reasoning. For this, we used a reasoning coding scheme  
developed in the context of organic reaction mechanisms  
(Kraft *et al.*, 2010; Christian and Talanquer, 2012) that includes  
15 model-based reasoning. These papers describe three main types  
of reasoning: rule-based, case-based, and model-based. Rule-  
based reasoning involves using a single rule or several rules  
(*e.g.*, octet rule), or relying on an algorithm. Case-based reasoning  
involves recalling a specific case (in this context a specific reac-  
20 tion) and matching it to the current problem, and can combine  
several rules specific to the case. Model-based reasoning involves  
generating and using a working mental model from chemical  
concepts and ideas, and dynamic mental models allow for  
simulation of the process during problem solving. In this context,  
25 model-based reasoning includes invoking general reaction types  
or mechanisms (*e.g.*,  $S_N1$ ). Mental models are developed as the  
result of consolidation of several cases and making general  
connections between them, which aligns with Lowe and Bou-  
cheix's phase 5 "Mental model consolidation".

## Methods

### Participants and setting

35 Prior to recruiting participants and conducting the study, ethics  
approval was granted by the institution's Research Ethics  
Board. Participants were students from a bilingual public  
research university in Canada who voluntarily participated in  
the study, and from two recruitment cohorts. Cohort A were  
40 Organic Chemistry II (OCII) students ( $n = 24$ ). Cohort B were  
students who were enrolled in or had completed either OCI or  
OCII ( $n = 26$ ). Five participants did not complete the full  
experiment and their data are excluded (final  $N = 45$ , 31 female†  
and 14 male†). Participants' ages ranged from 18–26 and  
45 reported their first language as English (32), French (8), Hindi  
(1), Spanish (1), Tagalog (1), or Ukrainian (1). The participants  
came from both English and French sections of the course,  
taught by different instructors, in which the organic chemistry  
sequence follows a patterns of mechanisms curriculum (Flynn  
and Ogilvie, 2015; Ogilvie *et al.*, 2017). Some sections of  
50 the courses are taught in a flipped format in which students watch  
videos or complete modules prior to coming to class (Flynn,  
2015). All participants gave written informed consent and were

† The demographic questionnaire asked "What is your gender?" with options  
55 "Male", "Female", "Prefer not to answer", and "These options do not apply to  
me, I identify as: \_".

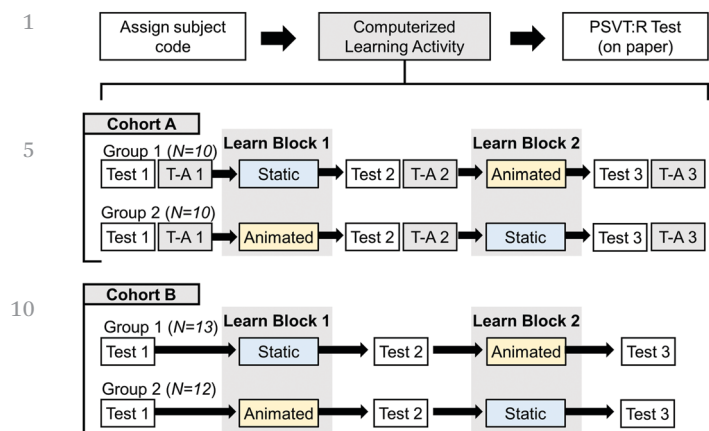


Fig. 1 Overview of the experiment. T-A = Think-Aloud. PSVT:R = Revised Purdue Spatial Visualization Test: Rotations.

compensated for their time. Pseudonyms are used in this article protect participants' identities.

## Experiment design and procedure

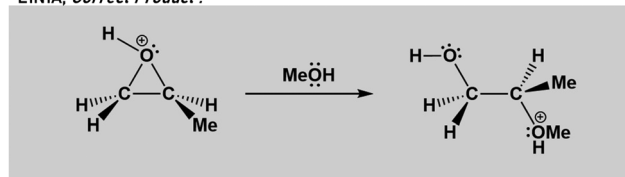
The experiment consisted of a computerized learning activity followed by a spatial ability test, the Revised PSVT:R (Maeda *et al.*, 2013). At the start of the interview each participant was assigned a successive code and placed into one of two groups (even code = Group 1, odd code = Group 2). Each group performed a series of consecutive tasks on the computer (Fig. 1) consisting of Tests and Learn Blocks, which were adapted from an earlier pilot study with in-depth qualitative interviews (Bongers *et al.*, 2019a, 2019b). The Learn 1 and Learn 2 task blocks were either static or animated, following a counterbalanced design. The Revised PSVT:R was then performed on paper. Cohort A also completed three short think-aloud (T-A) activities during the computerized learning activity, while cohort B did not. Cohort B was participating in an extension of the study that included brain scanning with electroencephalography (EEG), and the think-aloud interviews were removed due to time and operational constraints (Bongers *et al.*, 2019a, 2019b).

OpenSesame is an open-source program that was used to create and run the computerized experiment, which presents the stimuli and records accuracy and response times (Mathôt *et al.*, 2012). Examples of the simplified Organic Chemware<sup>®</sup> animations are provided in the Supporting Information. The complementary static images were created using ChemDraw software. All stimuli for the computer tasks (*e.g.*, test questions, static and animated representations) were displayed on a desktop computer screen with a grey background and black text and chemical structures. Research participants were seated approximately 20 inches from the bottom of the screen. Each stimulus was preceded by a black fixation cross on a grey screen (see Appendix, Fig. 11). A detailed description of the experimental paradigm for the Learn Blocks and Tests is provided below and in the Appendix.

## Tests

Tests 1–3 were all comprised of the same set of questions, shown to the participant in a random order. For each question,

Test Question: Determine if the product is Correct or Incorrect (Y/N).  
E1N1A, Correct Product.



E1N1A, Incorrect Product (wrong regiochemical outcome).

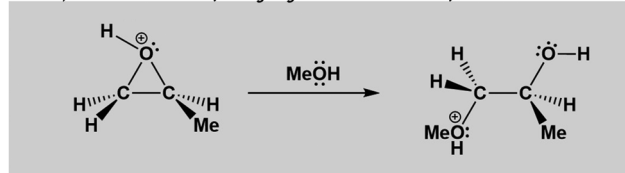


Fig. 2 Example of two test questions, one with correct and the other incorrect products. E1 = epoxide 1, N1 = nucleophile 1, A = acidic conditions. "Y/N" = Yes or No.

the participant was shown an epoxide opening reaction (Fig. 2), including starting materials and products; the task was to determine whether the product was correct or incorrect (*i.e.*, a binary response). The participant responded on the keyboard either with the left arrow (incorrect) or the right arrow (correct) using their right hand. The nucleophile was shown above the reaction arrow and no mechanistic (EPF) arrows were provided. There were 24 test questions representing 12 reaction types (see Appendix, Tables 6 and 7), prepared by the first author and evaluated by the second author. For each reaction type, there was one question with the correct product and one with the incorrect product. The test questions came from an earlier pilot study with in-depth qualitative interviews, which helped establish test validity (Bongers *et al.*, 2019a, 2019b). Response times (in seconds) and accuracy (values of 0 or 1) were recorded for each question, hereafter called "test response time" and "test accuracy", respectively. At the end of each test, participants were prompted on the screen to rate their confidence in their performance on a Likert scale of 1 (not at all confident) to 5 (completely confident).

## Learn blocks

In this study, we used simplified Organic Chemware<sup>®</sup> animations and complementary static images to explore how students learn and develop mental models of an organic reaction mechanism. Organic Chemware<sup>®</sup> animations were designed using Lewis and line structure symbolism that match the traditional representations (Nelson, "Organic ChemWare" 2018). These animations show lone pairs of electrons morphing into bonds in place of the EPF and show the spatial transition that the molecules undergo from starting materials to products. The full animation suite (not used in this study) also includes energy diagrams, molecular orbitals, and the static mechanistic scheme.

In the Learn Blocks (~15 min each), participants were shown either static or animated variations of the epoxide-opening reaction mechanism (Fig. 3). Only chemical structures were shown, with no text or audio explanations. There were 15 different variations of the reaction, each repeated 5 times in

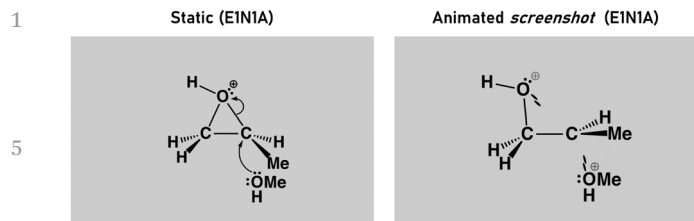


Fig. 3 Examples of static and animated stimuli from the learn blocks.

10 pre-set a randomized order, for a total of 75 trials in each block. All trials followed the same sequence and began with an image of the starting materials (0.8–1.2 s), followed by an image of the reaction mechanism (either static or animated, 3.2 s), followed by a final image of the product (1.0 s). Between trials (*i.e.*, the inter-trial interval) a fixation cross was presented on the screen.

15 There were two self-timed breaks in the block for participants to move or rest, after trials 25 and 50.

20 The epoxide-opening reaction mechanisms presented in the Learn blocks were comprised from combinations of four epoxides (E1–E4) and four nucleophiles (N1–N4) (see Appendix, Tables 6 and 7). The reactions were shown either under acidic or basic conditions. When the epoxide was unsymmetrical, the regiochemistry of the reaction/product depends on the presence of acid or base.

### Think-aloud interviews and qualitative analysis

25 After completing each test, participants in cohort A performed a short think-aloud interview, where they were shown one of the test questions (Fig. 4) and asked two different prompts: (1)

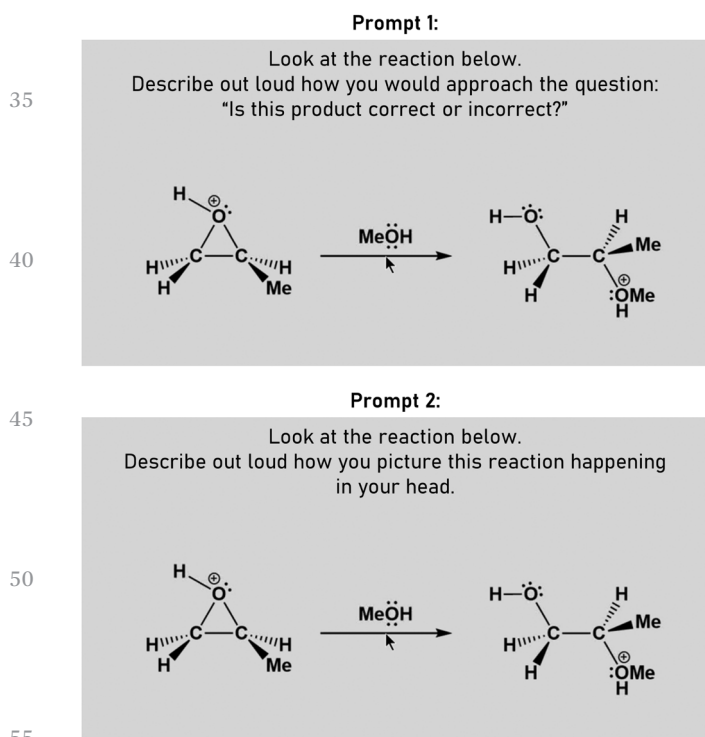


Fig. 4 Think-aloud interview prompts.

1 describe out loud how you would approach the question: is this product correct or incorrect? (2) Describe out loud how you picture this reaction happening in your head. These prompts were taken from our previous work on mental models when learning the epoxide-opening reaction mechanism (Bongers *et al.*, 2019a, 2019b). Participants were also specifically questioned about why the reaction occurs at one carbon site on the epoxide and not another. In total there were three think-aloud interview discussions for each participant which were audio-recorded then transcribed verbatim by the first and second authors. The data was then analysed using a qualitative lens as described below. One participant's audio data for Think-Aloud 3 was lost due to issues with the microphone.

10 Data from Prompt 1 were analyzed for the participants' claim (if they thought the product was correct or incorrect) and for how they made decisions to reach their claim (*i.e.*, their reasoning type). Reasoning types were coded as either rule-based, case-based, or model-based following prior literature (Kraft *et al.*, 2010; Christian and Talanquer, 2012). The goal of this analysis was to shed more light on each participant's test performance and how they were answering the test questions. This analysis would also reveal how the participant was visualizing the reaction mechanism (*e.g.*, used a dynamic mental model) to answer test questions. Data from the Prompt 2 were coded following the scheme outlined in our prior work (see Appendix, Table 5) to determine participants' working mental models throughout the activity. This coding identifies the participant's working mental models of the reaction as either static or dynamic, and further characterises these models in terms of their focus on symbolism, process, and particulate motion. Inter-rater reliability was tested by having 10% of the data (6 randomly chosen excerpts) coded by the last author, followed by discussion until full agreement was reached.

### Spatial ability test

15 The Revised Purdue Spatial Visualization Test: Rotations (PSVT:R) was used in this study (Bodner and Guay, 1997; Yoon, 2011). This questionnaire consists of 30 items and measures spatial visualization ability in three-dimensional mental rotation. Participants were given 25 minutes to complete the PSVT:R on paper, which was completed after the experiment to avoid fatigue during the chemistry tasks and to give the experimenter time to compile data. The PSVT:R instrument was found to be highly reliable (30 items;  $\alpha = 0.79$ ) in our participant sample ( $N = 45$ ).

## Results

20 The accuracy, response time, spatial ability, and confidence rating data from cohorts A ( $n = 20$ ) and B ( $n = 25$ ) were combined for quantitative analysis (Fig. 5). Equivalence of the cohorts was established using a one-way ANOVA which showed no significant differences between the cohorts for Test 1 accuracy ( $F(1,43) = 0.085$ ,  $p = 0.77$ ), Test 1 response time ( $F(1,43) = 0.45$ ,  $p = 0.51$ ), and PSVT:R scores ( $F(1,43) = 1.25$ ,  $p = 0.27$ ). After combining cohorts, the total  $N = 45$  (Group 1:  $n = 23$ , Group 2:  $n = 22$ ).



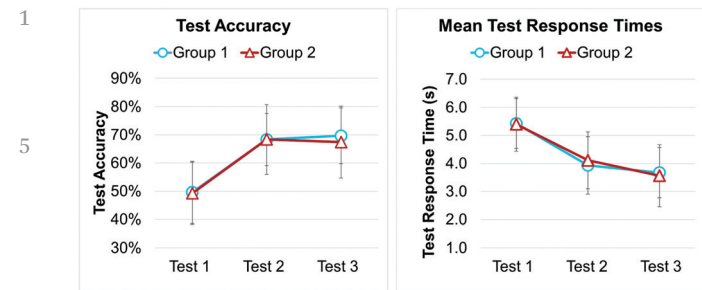


Fig. 5 Changes in accuracy and response time over Tests 1–3.

### No group differences in test accuracy and response times

The overall mean test accuracy increased from Test 1 ( $M = 50\%$ ,  $SD = 11\%$ ) to Test 2 ( $M = 69\%$ ,  $SD = 11\%$ ) and Test 3 ( $M = 69\%$ ,  $SD = 11\%$ ),  $N = 45$ . These tests were all the same and participants received no feedback. The mean accuracy for Test 1 was equal to chance, which suggested guessing on this binary response test. Accuracy did not change on average between Tests 2 and 3, which suggested there was no test-enhanced learning effect. The overall ( $N = 45$ ) mean test response time decreased from Test 1 ( $M = 5.43$  s,  $SD = 0.92$  s) to Tests 2 ( $M = 4.02$  s,  $SD = 1.02$  s) and Test 3 ( $M = 3.61$  s,  $SD = 0.99$  s). While these variables changed from Tests 1–3, we found high positive within-subject correlations for participants' accuracies in Tests 1–3 (Pearson  $R = 0.402$ – $0.558$ ,  $p < 0.01$ ) and also response times (Pearson  $R = 0.387$ – $0.784$ ,  $p < 0.01$ ).

Changes in measures of test accuracy and response time over time (Fig. 5) were explored using a mixed repeated-measures ANOVA. The independent variable group (1 or 2) was included in the analysis to test for effects of learning from static *versus* animated representations. We found a significant main effect of time on test accuracy ( $F(2,86) = 88.2$ ,  $p < 0.001$ ,  $\eta_p^2 = 0.672$ ) and test response time (Greenhouse–Geisser corrected:  $F(1.57,67.54) = 100.2$ ,  $p < 0.001$ ,  $\eta_p^2 = 0.700$ ). Bonferroni corrected pairwise comparisons showed an increase in accuracy ( $MD = 19\%$ ,  $p < 0.001$ ) and decrease in response time ( $MD = -1.41$  s,  $p < 0.001$ ) from Test 1 to Test 2, but no differences between Tests 2 and 3. However, there was no effect of group on test accuracy ( $F(1,43) = 0.033$ ,  $p = 0.86$ ) or response time ( $F(1,43) = 0.073$ ,  $p = 0.79$ ), or the time\*group interaction (accuracy  $p > 0.9$ , response time  $p > 0.3$ ). Learning gains were calculated as the normalized changes in accuracy between Tests 1 and 2 (mean  $LG1 = 38\%$ ,  $N = 45$ ) and Tests 2 and 3 (mean  $LG2 = -5\%$ ,  $N = 45$ ) as shown in eqn (1) and (2).

$$LG1 = \frac{\text{Accuracy}_{\text{Test2}} - \text{Accuracy}_{\text{Test1}}}{100\% - \text{Accuracy}_{\text{Test1}}} \quad (1)$$

$$LG2 = \frac{\text{Accuracy}_{\text{Test3}} - \text{Accuracy}_{\text{Test2}}}{100\% - \text{Accuracy}_{\text{Test2}}} \quad (2)$$

### Test confidence ratings increased for both groups

Non-parametric tests were used to analyse test confidence ratings (Likert-scale) for the three tests and across groups (Fig. 6). The average test confidence ratings for all participants

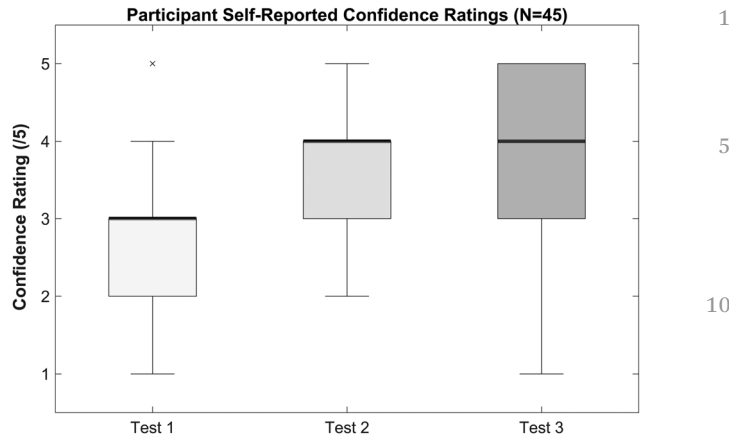


Fig. 6 Boxplot of confidence ratings ( $N = 45$ ) showing median in a bold line.

( $N = 45$ ) followed the same trend as test accuracy, increasing from Test 1 ( $M = 2.6$ ,  $Mdn = 3$ ,  $SD = 0.87$ ) to Test 2 ( $M = 3.7$ ,  $Mdn = 4$ ,  $SD = 0.84$ ) and staying the same in Test 3 ( $M = 3.9$ ,  $Mdn = 4$ ,  $SD = 0.94$ ), and a Friedman test showed these differences were significant ( $\chi^2(2) = 66.9$ ,  $p < 0.001$ ). However, the median confidence rating for Groups 1 ( $n = 23$ ) and 2 ( $n = 22$ ) was the same for each of the tests (Mann–Whitney  $U$  test,  $p > 0.6$  for Tests 1–3).

### Spatial ability influenced learning

We looked at how spatial ability was related to measures of test accuracy and response times (Fig. 7). Spatial ability (as measured by PSVT:R score) was found to be normally distributed in the sample (Fig. 7A) and comparable to similar student populations (Yang *et al.*, 2003; Maeda *et al.*, 2013).

Spatial ability was positively correlated to test accuracy with Pearson  $R$  values from 0.296–0.486 (Fig. 7B). The participants' average test response time was also positively correlated to spatial ability scores in Test 1, but not for Tests 2 and 3 (Fig. 7B). A Spearman correlation analysis found no relationship between test confidence ratings and spatial ability.

We also investigated the effect of participants' spatial ability on the observed increase in test accuracies between Test 1 and Test 2 (LG1). Visual inspection of the data suggested that Group (1 = static, 2 = animated) may have interacted with spatial ability to influence LG1 (Fig. 8).

Backward multiple regression was conducted to see how group, spatial ability (PSVT:R score), and the interaction between these variables influenced learning gains (Table 1). The categorical variable Group was recoded (Group 1 = -1, Group 2 = 1) to correct for collinearity of the interaction variable. We first entered all variables and used backward elimination method to exclude variables (criterion: probability of  $F$ -to-remove  $\geq 0.1$ ) from the model as shown in Table 1. The final model showed that only spatial ability scores were a significant predictor of learning gains ( $F(1, 43) = 7.51$ ,  $p < 0.01$ ,  $R^2 = 0.149$ ) with a small effect size.

### Claims and reasoning changed during the activity (Prompt 1)

For Prompt 1 (Fig. 4), we investigated the participants' claims about whether the product shown was correct or incorrect (the

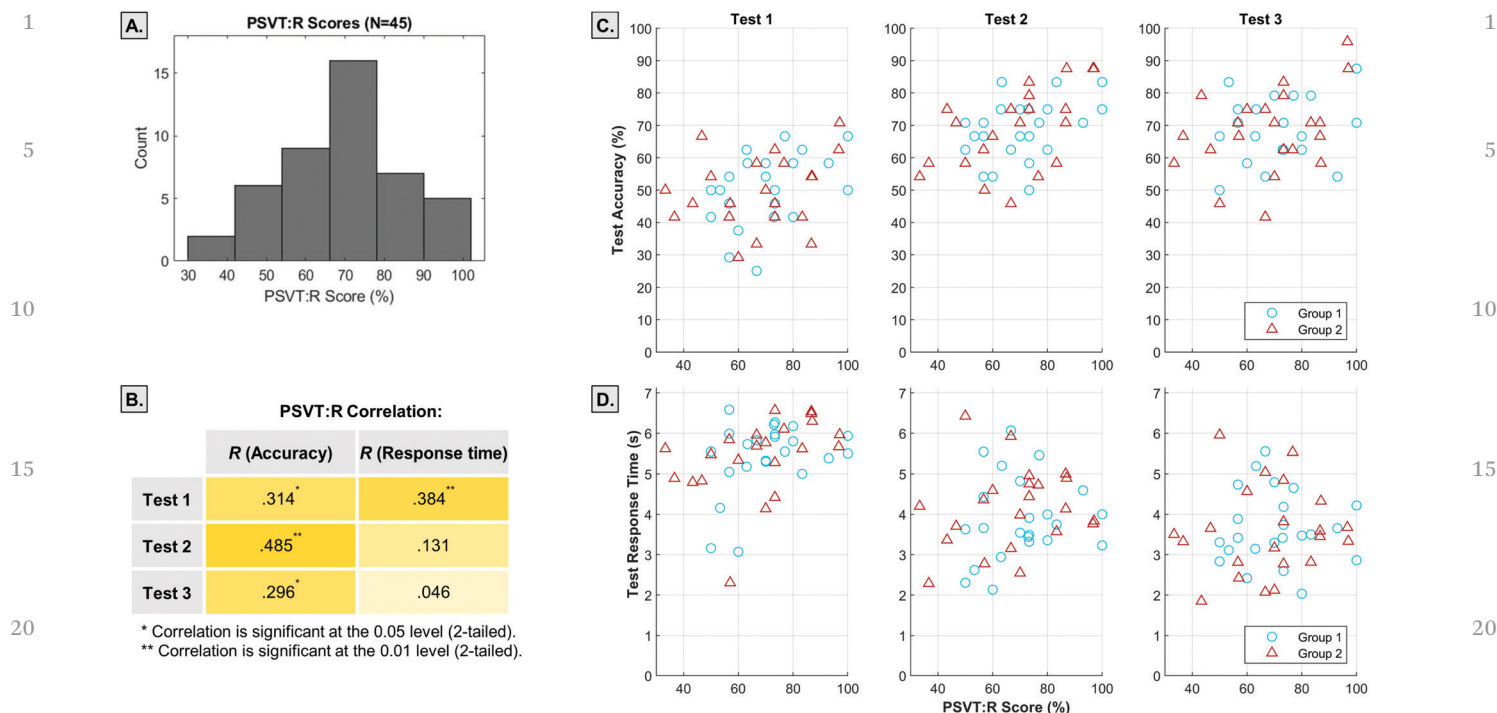


Fig. 7 (A) PSVT:R score distribution ( $N = 45$ ) and (B) Pearson correlations to test accuracy and test response time. Scatter plots showing (C) test accuracy (%) and (D) test response time (s) as related to spatial ability (PSVT:R score, %).

25

25

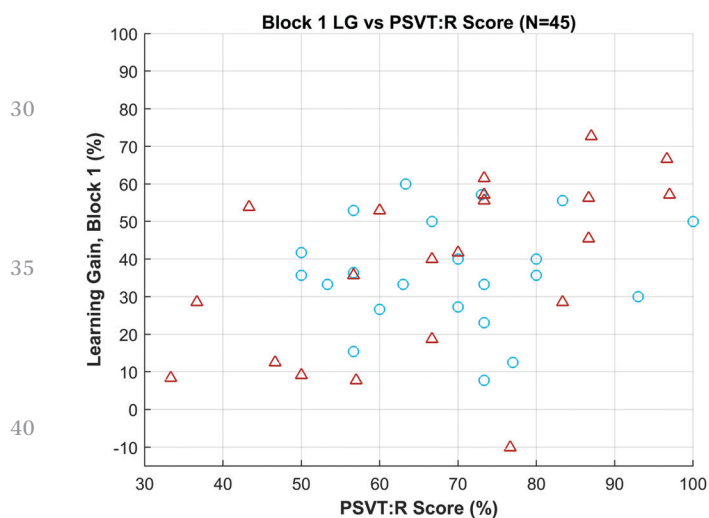


Fig. 8 Scatter plot of Block 1 learning gain (LG1) and spatial ability (PSVT:R score).

45

product shown was correct). The trends in claims were similar to the observed trends in test accuracy (Fig. 9), with Think-Aloud 1 being at chance (50% accuracy) and increasing accuracy for Think-Aloud 2 (75%) and Think-Aloud 3 (89%). No differences were observed between groups.

Participants' reasoning types changed throughout the learning activity, beginning with more occurrences of model-based and ending with more rules-based reasoning (Fig. 9). Participants often used multiple reasoning types in their responses depending on which feature in the prompt was being discussed. There were no differences in reasoning types between Groups 1 and 2. Quotes from participants during reasoning for Prompt 1 are described below and in Table 5.

In Think-Aloud 1, 16 out of 20 participants attempted to use the reaction mechanism to determine if the product shown in the prompt was correct. This is an example of model-based reasoning: these participants simulated the mechanism in their mind and checked if their simulated product matched the one on the screen. These participants' utterings showed that they

30

35

40

45

Table 1 Multiple regression models for spatial ability, group, and learning gains<sup>a</sup>

Predictor variable	Model 1			Model 2			Model 3		
	$B$	SE $B$	$\beta$	$B$	SE $B$	$\beta$	$B$	SE $B$	$\beta$
Group	0.165	0.116	-0.898						
Group $\times$ PSVT:R	0.257	0.163	0.992	0.031	0.037	0.119			
PSVT:R	0.393	0.163	0.345*	0.444	0.161	0.390**	0.439	0.160	0.385**
$R^2$		0.202*			0.163*			0.149**	
$F$ for change in $R^2$		3.459*			2.020			0.707	

55

55

<sup>a</sup> SE = standard error,  $B$  = unstandardized coefficient beta,  $\beta$  = standardized coefficient beta,  $R$  = correlation coefficient. \*  $p < 0.05$ , \*\*  $p < 0.01$ .

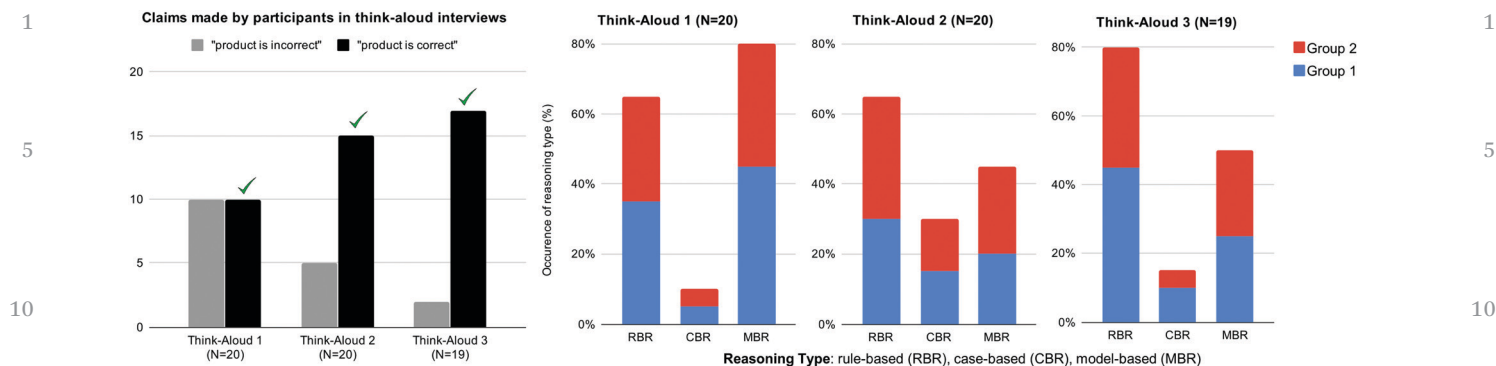


Fig. 9 Claims made by participants and occurrence of reasoning types in think-aloud interviews (Prompt 1). Multiple reasoning types may have been observed for each participant (sum is greater than 100%).

used a dynamic mental model of the process in their answer (Quote 1). Of the participants who thought the product shown was incorrect in Think-Aloud 1 ( $N = 10$ ), 6 reasoned that the reaction would either occur at the less hindered carbon site (Quote 2) and the other 4 thought it may not occur at all (*i.e.*, that the reactants would undergo a different reaction, Quote 3). Many of these participants also used rule-based reasoning while explaining their claim further, especially when considering if the formal charges were correct in the product.

In Think-Aloud 2 and 3, many participants changed their reasoning after seeing new information in the learn blocks (Fig. 9). Some participants appeared to switch from using mental models of the reaction mechanism (20% decrease in the occurrence of MBR) to using simple heuristics to determine of the product was correct (RBR, Quote 4) or remembered specific cases from the learning activity (CBR, Quote 5). Most participants who previously thought the product was incorrect due to sterics or plausibility now claimed the product was correct based on these rules or cases, but expressed not knowing why the product was correct; in essence, these participants did not use their prior knowledge or mental models of reaction mechanisms in their answer (Quote 6). These heuristics included some based in chemistry concepts (octet rule, conservation of charge) and also some based solely in patterns in the symbols (*e.g.*, if this H points to the left, the new group adds to the carbon on the right) (Quote 7, Quote 8).

However, some participants maintained the use of their mental model of the reaction mechanism in their reasoning and tried to incorporate the new information into their model. These participants envisioned multiple possible mechanisms and products, and used MBR to reason about which mechanism would lead to the major product (Quote 9) (Table 2).

### Working mental models changed after learning from animations (Prompt 2)

For Prompt 2 (Fig. 4), we used the working mental models coding scheme (Appendix, Table 4) to determine the participants' working mental model of how they pictured the reaction happening throughout the experiment. In Think-Aloud 1, prior to the learning blocks, participants' working mental models were similar between Groups 1 and 2 (Fig. 10) and were almost

evenly distributed between the three types: static, dynamic process, and dynamic particles in motion. Some participants' working mental models changed after the learning activities (Fig. 10). After the animated learning activity, there was a large increase in participants working with a dynamic particles in motion mental model (Group 1 = +30%, Group 2 = +15%). After the static learning activity, Group 1 showed a small increase in static (+5%) and dynamic process (+5%) working mental models, while Group 2 showed a further increase in using the dynamic particles in motion mental model (+12%). Representative quotes for the three types of working mental models from participants in the study are provided in Table 3.

In summary, the participants demonstrated using generally dynamic working mental models to visualize the reaction mechanism (Prompt 2) before and after learning blocks. The type of working mental model used by some participants for visualization changed throughout the learning activity, especially after animated learning blocks. These participants expanded their mental models and built on links to prior knowledge, which reflects progression to phases 3 and 4 in Lowe and Boucheix's framework:

Participant 9, T-A 1, Prompt 2: "So for this one I—I would... I would imagine the bond breaking, um, between the charged oxygen and the carbon on the right, and those electrons going towards the oxygen, um, and then... yeah. The molecule over the reactant arrow would come in, um, from the bottom to where the carbocation and the other atoms attached to it... sorry [laughing]. Um, it would just be with—it would be moved up so they would keep the same orientation I think, they just move upwards."

Then, after learning from animations:

Participant 9, T-A 2, Prompt 2: "So again I guess I look at the product first and then the starting material and see um, if I set it up with the molecule over the arrow um, beneath the bottom right carbon, the bond between that carbon and the um, charged oxygen on top would break and it would—it would smoothly kind of arrange itself into that final product while the bond is forming um, and the bond is breaking so yeah, that's what I would see and then just to confirm that it is the right product I'd check the stereochemistry if the groups attached to the carbon."

Interviewer: "When you say smoothly can you tell me a bit more about, um, what that means to you?"

1 Table 2 Example quotes showing reasoning types during the different stages of the interview 1

Quote #	Example quotes from participants in response to Prompt 1	Reasoning
1	Participant 13, T-A 1: "First it would—the oxygen and um, on the methyl would attack hydrogen and the oxygen give the lone pair um, next the um. . . the carbon with the two Hs attack the methanol, um, ion and the OH would be pushed to the right side and then the carbon on the left would be left as a cation and um, another methanol compound will attack the cation and, yeah it's not a correct product."	Model-based
5	Participant 22, T-A 1: ". . . So the MOH [sic], the O is going to go to. . . it's gonna attack one of the carbons. . . and. . . [long pause] I would say that the final product is not correct, because I think you will, that methanol will be attacking the carbon that is the least substituted, so that [it's] not stéréochimiquement [stereochemically], like encumberated [sic], I honestly don't know in English, encombrer [encumbered]."	Model-based
10	Participant 16, T-A 1: "Um, well you have to kinda like evaluate the thing-a-majig [laughs], the solvent but the two like products like the two reactants that you have like MeOH is like a—I think it's a weak base, yeah 'cause [NaOMe] is a strong base, yeah, I think it's a weak base and the weak base—and then what's—the type of reaction that would be happening through this?"	Model-based
15	Interviewer: "Ok, um, so you described a bit of a different mechanism than last time, can you tell me about that?" Participant 13, T-A. 2: "Because um, like all the videos like before it kinda showed the trend of um, what is it like the solution or whatever, the substrate, um, attacks the carbon that has um, if it has more um, alkyl groups and if it will have like net negative or positive charge um, it will tend to favour the carbon with more alkyl groups".	Rule-based
20	Participant 12, T-A 2: "Um, I do believe that this product is correct um, because the um, from like the previous when we saw a bunch of mechanisms it kinda stands true with that and the lone pairs on the methanol attack the right most carbon methyl group, and the stereochemistry of the methyl group and hydrogen groups stay the same as they did in those, and the bond between the oxygen and right carbon breaks, and um, removing [sic] the charge on the hydroxyl group."	Case-based
25	Participant 22, T-A. 2: "Uh, well there must be a chemical reason because if it's MeO with a negative [charge], I know like if there's no H in the reactive [sic], if it's a charged molecule it's gonna be attacking the least substituted [sic] one. It's gonna be attacking the carbon with the two Hs on it, so. . . um. . . it's not, it's a . . . not a charged molecule. . . uh. . . I don't know why!"	Rule-based
30	Participant 6, T-A 3: "Because, that new group shows up on the right side. So, in my mind, that bond from the right side connecting the carbon to the oxygen, that's the one that's going to break to create a positive carbon and attach that there. [pause] Yeah. If it was on the left side, the top oxygen would be connected to the right ox—the right carbon, instead of the left carbon. Yeah."	Rule-based
	Participant 12, T-A 3: ". . . the way I'm just visualizing is kind of with a way the hydrogen on the hydroxyl is, it's kinda to the left so it kinda just like—I know that it doesn't sound right but like—the methanol is to the right of the starting material so it makes sense if it attacked on the right side and then pushed the hydroxyl to the left. That's kind of like what makes sense to me."	Rule-based
	Participant 18, T-A 2: "I can still see the elimination product. Um, but, like this is a viable reaction. The only debate I have is if the [epoxide] oxygen will leave first or if the methox—the methanol would attack first, but since it's so weak, it must. . . the oxygen has to leave before the methanol can attack."	Model-based

Participant: "Um, I guess I'm just imagining the, um, the videos of the reactions that happened before so, if it's happening smoothly it means that there's no, um, geometrical hindrance I guess to that happening, it just – it can proceed, um, and nothing has to spin or flip, you can just go ahead."

## 40 Discussion

### 1. How does learning a reaction mechanism from static or animated visualizations affect measures of learners' test accuracy, response times, and confidence?

45 We observed no differences between static and animated learning conditions for participants' test accuracies, test question response times, or confidence ratings. This finding is different from many prior reports (Stieff and Wilensky, 2003; Tasker and Dalton, 2006; Kelly and Jones, 2008; Aldahmash and Abraham, 2009; Al-Balushi and Al-Hajri, 2014; Kelly and Akaygun, 2016), although negative results often go unreported and could influence the publication of studies where animations are found less or equally effective as static representations (Fanelli, 2010). We found it encouraging that the students appeared to learn the reaction (or more accurately, how to identify its products) from the animation in the absence of

explanations and without the traditional EPF. We suspect that answering questions on the tests, which contained only static representations, did not require skills such as mental rotation of molecules or simulation of the reaction which are reported to be gained from viewing animations (Stieff *et al.*, 2011). This consideration of the assessment could help explain why there are conflicting reports of the benefits and limitations of learning from animations (Höffler and Leutner, 2007; Stieff, 2011a). Indeed, our findings support the idea of a local, (rather than global) model for visualization in chemistry where the characteristics of the individual learner, including spatial ability and reasoning strategy, are important factors in performance (Stieff, 2004, 2011b; Stieff and Raje, 2010; Stieff *et al.*, 2014). We then turned to qualitative data from the think-aloud interviews to shed some light on why students' performance on the tests was not influenced by visualization type.

The think-aloud interviews revealed that in the absence of explanations, some participants preferred to use simple rules and patterns for the test rather than working with their dynamic mental models of the reaction mechanism. Qualitative analysis of Prompt 1 also showed that participants' reasoning strategies were more models-based before the learning blocks, where many participants worked through or visualized the reaction mechanism to reach their answers. After Learn Block

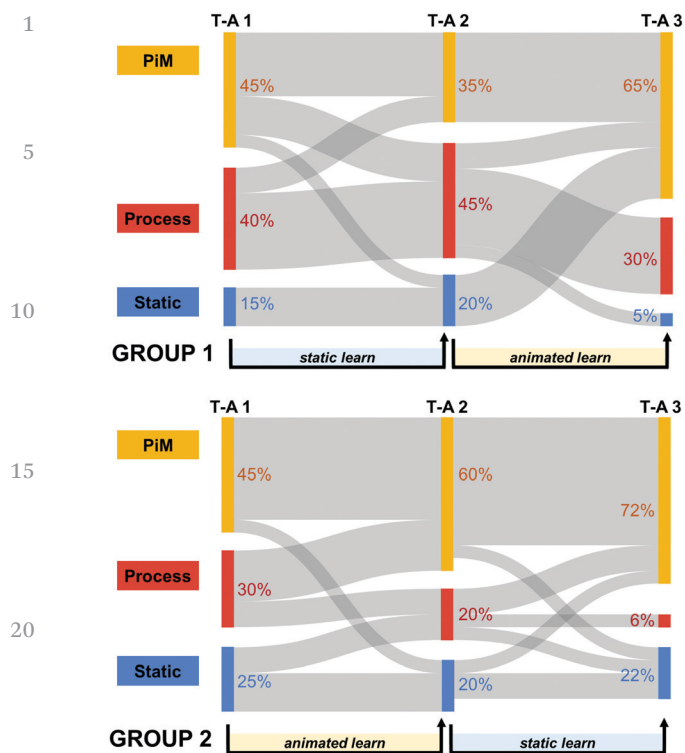


Fig. 10 Changes in working mental models (from Prompt 2), three types: static (blue), dynamic process (red), and dynamic particles in motion (PiM, yellow).

1, many participants switched their reasoning strategy and the analysis showed an increase in rule- and case-based reasoning, which is more commonly reported in the literature (Bhattacharyya and Bodner, 2005; Kraft *et al.*, 2010; Christian

and Talanquer, 2012; Weinrich and Talanquer, 2016; Caspari *et al.*, 2018; Moreira *et al.*, 2019). Participants may have switched to rule- and case-based reasoning strategies because these strategies are quick and do not require mental rotations or critical analysis of the question. The short window to answer each question thus led participants to use the learning blocks to pick up on useful patterns in the symbols to answer the tests, sometimes without understanding or considering the chemical reasons for these rules. This use of heuristics corresponds with the decrease in question response times after the learning blocks (Tests 2 and 3). This change in strategy by the participants validates the finding of no difference in test accuracy between learning from static or animated representations.

Considering model-based reasoning as the “highest level” of reasoning (model-based > case-based > rule-based) is justified by Lowe and Boucheix’s framework but is not necessarily linked to expertise or sophistication of the reasoner. For example, phase 1 involves a mental model with specific and salient information that would be used in rule-based reasoning, while a consolidated mental model (phase 5) is one that can be applied to related systems in model-based reasoning. However, there are many cases where rule- or case-based reasoning are suitable and functional reasoning types. In this study, we found some evidence that the participants’ reasoning types were related to what feature being reasoned about (*i.e.*, rule-based reasoning about formal charge, but model-based reasoning about product stereochemistry), which will be followed up on in a future research.

The tests we used in this study were unique and experimental, but still align in many ways with students’ test taking experiences including quizzes online, short times to answer questions during exams. This study lends more support to the

Table 3 Examples quotes from participants who described their working mental models (WMM) for how they pictured the reaction happening

WMM	Example quotes from participants in response to Prompt 2
Static	Participant 16, T-A 1: “I’m trying to visualize my chart in my head um, yeah I think an $S_N2$ reaction requires strong nucleophile and then MeOH is strong nucleophile—I just did a midterm, I should remember um, I feel like MeOH is like a decent nucleophile that’s just kind of like in my head like I have that there like there’s no actual logic to it, so to understand the [inaudible] step—that’s another thing um, so I think it’s like a decent enough nucleophile.” Participant 22, T-A 2: “I’m not entirely sure if this is correct. . . but in the last section where it kind of just showed you mechanisms, typically like I noticed that the direction in which the hydrogen atom was pointing, the one that’s bonded to the [epoxide] oxygen already, like on the top in the reagent, um, dictated whether it would go to the left or the right. So I feel like the oxygen, the carbon–oxygen bond on the right, would break, and it would move to the left, just because that’s where the hydrogen is pointing, or directed, I don’t know if that has any actual scientific, um, use or. . . importance, but, I don’t know how, I just looked at it and saw that.”
Dynamic process	Participant 4, T-A 2: “Uh, so the oxygen on the methanol would—the electrons would take, uh, it would be an acid–base reaction with the hydrogen on the oxygen on the epoxide, so the bond between the hydrogen and the oxygen on the epoxide would break, the electrons go back to the hydrogen on the epoxide, and the hydrogen would then bond to the oxygen on the methanol, uh, creating a charged methanol.” Participant 17, T-A 2: “So this is the same um, mechanism and product is still correct um, the positive charge is on the methoxy group on the oxygen and of the nucleophile and there’s a new lone pair that came from the bond that broke between the oxygen and the alpha carbon um, forming the hydroxyl group on the left again on the least substituted carbon.”
Dynamic particles in motion	Participant 13, T-A 2: “So in like, um, they would undergo like random motion and the oxygen with the lone pairs will collide with the carbon, um, on the right and since the ring is very, um, hindered it will want to move on and uh, kick out the electron that is bounded to the oxygen and then the carbon will like get rid of that, um, electron and then it will pick the bond with the oxygen atom on the methanol.” Participant 9, T-A 3: Um, so again I guess I just—I pictured um, like I mentally put the molecule up in the arrow at the bottom right underneath and I see it approaching that carbon and it’s forming a bond while that carbon um, and the um, charged oxygen are breaking the bond at the same time so it’s all happening at the same time um, and then guess the electrons either flow to form a bond or break a bond.

1 Table 4 Details of Lowe's 2008 framework for learning from animations 1

Phase	Description of phase (Lowe and Boucheix, 2008)
1. Localized perceptual exploration	Initial "bottom-up" perception and segmentation of dynamic information in the animation into event units. This tracking of localized events over space and time depends on perceptual salience of the information in the animation to the learner.
2. Regional structure formation	Linking event units based on features like proximity or behaviour. Coordination of multiple events and their interactions based on general prior knowledge.
3. Global characterization	Building on regional structure formation, the learner adds details ( <i>e.g.</i> , onset, direction, and magnitude) and extends their model of the system over time to establish causal links within the process. These causal links may be mischaracterized if the learner does not have the appropriate domain-specific knowledge.
4. Functional differentiation	The learner interprets the events and causality within the animation in terms of the purpose or function of the system. This involves recruitment of prior knowledge and "top-down (away from perception)" processing. This phase could be difficult for domain novices or learners without the necessary specialize knowledge, but is essential for building a robust mental model of the system.
5. Mental model consolidation	The learner consolidates information in their mental models for application to a wide variety of circumstances and range of performances. This phase is key to learning across disciplines and to building a high quality mental model that applies to other situations and systems. The nature of animations could limit mental model consolidation, since only one specific sequence of events is presented in the animation.

need for redesigned assessments, since some types of tests can be simply measuring students' ability to use on patterns and rules in the symbols. Assessment items for organic reaction mechanisms that are designed around learning outcomes such as "draw the products" and "draw the arrows" do not measure students' mental models of reactivity (Flynn and Featherstone, 2017). While heuristic reasoning is an important skill for chemists (Graulich *et al.*, 2010; McClary and Talanquer, 2011), educators must try to measure other abilities on their assessments and help students identify their own strategies and choose which are most appropriate for a given context (Stieff and Raje, 2010).

## 2. What is the influence of the learner's spatial ability on the above measures?

Participants with higher spatial ability, as measured by the PSVT:R, performed better on the tests. This correlation is in agreement with some prior work in organic chemistry education (Bodner and McMillen, 1986; Carter *et al.*, 1987; Yang *et al.*, 2003; Supasorn *et al.*, 2008). In this experiment, this correlation is likely related to the need for quick responses during the tests. Spatial ability was also correlated to Test 1 response time, but not Test 2 or 3. These findings, along with the reasoning strategies, suggest that the participants were taking time to mentally visualize the reaction mechanism (which requires spatial ability) to answer the questions in Test 1 but not on subsequent tests. Analogously, more participants described that they visualized the reaction mechanism while demonstrating model-based reasoning in Think-Aloud 1 but switched to using heuristics in Think-Aloud 2–3. Stieff *et al.* (2012) similarly found that students used multiple problem-solving strategies (*e.g.*, spatial-imagistic, heuristic) and that students of lower spatial ability were more likely to use or switch to heuristics and algorithmic strategies (Hegarty *et al.*, 2013).

To explore the influence of spatial ability on learning from different representations, we used normalized learning gains as a measure of learning that occurred during the learning task blocks. Only the learning gains in Learn Block 1 (LG1 = 38%) were relevant for this analysis, since the test accuracies did not

change after Learn Block 2 (LG2 ~ 0%). We predicted that learning gains would be influenced by spatial ability and that this relationship may be stronger for participants who had learned from animated representations. This hypothesis was informed by prior work (Pribyl and Bodner, 1987; Aldahmash and Abraham, 2009; Höffler, 2010) and Lowe and Boucheix's (2008) framework for learning from animations, where the upper phases 3 and 4 (global characterization and functional differentiation) could require spatial coordination for understanding of the system. Backward multiple regression showed that spatial ability was a moderate predictor of learning gains, regardless of representation type. This finding is in agreement with a study into learning from electrochemistry animations (Yang *et al.*, 2003), where spatial abilities related to performance in a post-test, there was no interaction effect for spatial ability and animated *vs.* static learning conditions.

As discussed above, analysis of qualitative data from the think-aloud interviews demonstrated that the tests were a limited measurement of learning and knowledge of the reaction mechanism. The influence of an individual learner's spatial ability on their mental models and modelling skills in organic chemistry remains underexplored. Prior work has shown that the adoption of visual/spatial strategies is challenging for organic chemistry students (Hegarty *et al.*, 2013; Vlacholia *et al.*, 2017), and possibly more so for students with lower spatial ability (Stieff *et al.*, 2012). Students with high mental modelling abilities are able to not only construct but apply their models to problem solving, and incorporate new content knowledge (Wang and Barrow, 2011).

## 3. How does learning a reaction mechanism from static or animated visualizations affect learners' working mental models of the reaction?

While static representations of organic reaction mechanisms convey information with EPF, the motion in an animation conveys the same information with fewer symbols. However, animated representations may be less accessible to students with lower spatial ability. We found that after learning from animations, participants in both groups used more dynamic

1 and transitional language to explain the reaction mechanism  
they visualized, using terms like “flow” for electron transfer or  
describing the molecules moving in the reaction. Moreover, the  
mental models described for visualizing the reaction (Prompt  
2) were found to be clearly different than those used to reason  
(Prompt 1) in Tests 2–3. This divergence reflects the other side  
of the findings above, that even though participants’ visualiza-  
tion mental models were changing and developing, many were  
not using visualization strategies during the tests. Studies show  
that visualization (or imagistic) strategies are not necessary for  
all types of chemistry problem solving and may be difficult for  
learners with low spatial ability (Stieff, 2011b; Stieff *et al.*, 2012,  
2014). Indeed, experts often use algorithms to solve spatial  
tasks (Stieff and Raje, 2010).

15 In Think-Aloud 1, many participants leveraged prior knowl-  
edge of reaction processes or dynamics to visualize the reac-  
tion, which aligns with prior work in the same curriculum  
(Galloway *et al.*, 2017; Bongers *et al.*, 2019a, 2019b). In Think-  
Aloud 1, these participants also used these mental models to  
reason and determine if the products were correct. Learning  
from animations allowed these participants to build on their  
dynamic mental models to include more sophisticated transi-  
tional features. Within Lowe and Boucheix’s framework, these  
participants were advancing to phases 3 and 4 by extending  
their mental models to the whole system and finding and  
causal links using their prior knowledge (Quotes 10 and 11).

Participants who began with and maintained a static mental  
model of the reaction (using rules and patterns without process)  
were using basic level information obtained from the animations,  
staying in Lowe and Boucheix’s phase I. This information was the  
same as shown in the traditional representations, and these  
participants’ discussions in the think-aloud interviews focused on  
contextually salient symbols and patterns, such as formal charges  
and lone pairs of electrons (Lowe, 1999; Kelly *et al.*, 2017).

## Limitations

40 This study used two participant cohorts with different sampling,  
where all participants had in common that they were or had been  
enrolled in OCII at the university of this study. Cohort A participated  
in the experiment with think-aloud interviews, while Cohort B  
participated in the same study without the interviews. The interview  
questions may have influenced the learning of participants in Cohort  
A; they also had more time to consider the material. However, those  
questions and time did not seem to affect test accuracies and  
response times, as both cohorts were found to be equivalent.

50 The design of this study required the materials (*i.e.*, stimuli) in  
the learning activities and tests to be simple as possible, therefore  
no explanations were provided. Also, to show the participants  
multiple variations of the epoxide-opening reaction mechanism,  
multiple stimuli were presented in sequence for a short time on  
this screen. For the same reasons, the tests required a time limit for  
each question, which reduced the opportunity for chance successes  
in the binary response. This learning environment almost definitely  
inclined participants to shift towards quick heuristic styles of

reasoning answering the tests, and likely influenced how they  
viewed the static and animated representations. This learning  
environment may seem unusual, but in many ways (unfortunately)  
aligns with typical learning environments: students must take in  
information and make decisions at rapid speed while watching  
online videos, flipping through the textbook, or following along to  
slides in class.

## Conclusions

10 This study supports the use of Organic Chemware<sup>®</sup> animations to  
help students build on their prior knowledge of reaction processes  
and molecular dynamics to develop dynamic mental models of  
reaction mechanisms. However, we found a gap in students’ ability  
to transfer these visualization mental models to reasoning in time-  
limited test environments. This transfer of modelling skills to  
reasoning aligns to Lowe and Boucheix’s phase 5, where mental  
model consolidation allows the learner to apply their knowledge to  
new situations and systems. From these findings and our prior  
work (Bongers *et al.*, 2019a, 2019b), we recommend a classroom  
focus on practicing how to use dynamic models of reaction  
mechanisms to reason about a process to bridge this gap.

15 Before learning about the reaction and its mechanism in our  
study, the participants worked more with prior knowledge of  
reaction mechanisms in general to answer test questions, using  
their own mental models for model-based reasoning. When these  
mental models were dynamic and simulative, participants could  
envision multiple possible outcomes for the reaction, and consid-  
ered steric hindrance, nucleophilicity, and electrophilicity in their  
answers. Considering multiple factors in a system and their prob-  
able influences is important to learning and reasoning about  
chemistry, and something should be nurtured in students. How-  
ever, simulative or visualization strategies are not always necessary  
for the task. After switching to rule- and case-based reasoning  
strategies participants’ confidence and accuracy of their claims  
improved, but they no longer considered reactivity beyond what  
was shown on the page. This study shows the advantage and costs  
using rules and heuristics for reasoning in chemistry, and the  
limits of assessments in evaluating imagistic reasoning.

## Implications

45 Using models and developing mental models are invaluable  
skills for students to learn, and our findings imply that organic  
chemistry animations can be used to bolster students’ mental  
models to include dynamic interactions of molecules and  
particles. Educators can use students’ natural tendencies to  
use deeper explanations (Galloway *et al.*, 2017) but need to  
demonstrate the value of deeper, causal explanations in  
instruction and assessment, lest students switch to less scien-  
tifically meaningful methods of answers to simply get to the  
answer (Bodé *et al.*, 2019; Crandell *et al.*, 2019). This demon-  
stration would help students identify when to use model-based  
reasoning in problem-solving or in assessments, and situations  
where rule- or case-based reasoning strategies are suitable.

Educators should also be aware that spatial ability is related to students' performance in tests, and time-limited environments may drive students towards using memorized rules. Assessments that are meant to evaluate model-based, spatial, or imagistic reasoning in chemistry must allow time for these tools to be used by students with varying spatial abilities and confidence.

## Conflicts of interest

There are no conflicts to declare.

## Appendix

### Theoretical frameworks

#### Overall experiment and design

The overall experimental design and epoxides (E1–E4) and nucleophiles (N1–N4) used to make the experimental material are provided in Fig. 11.

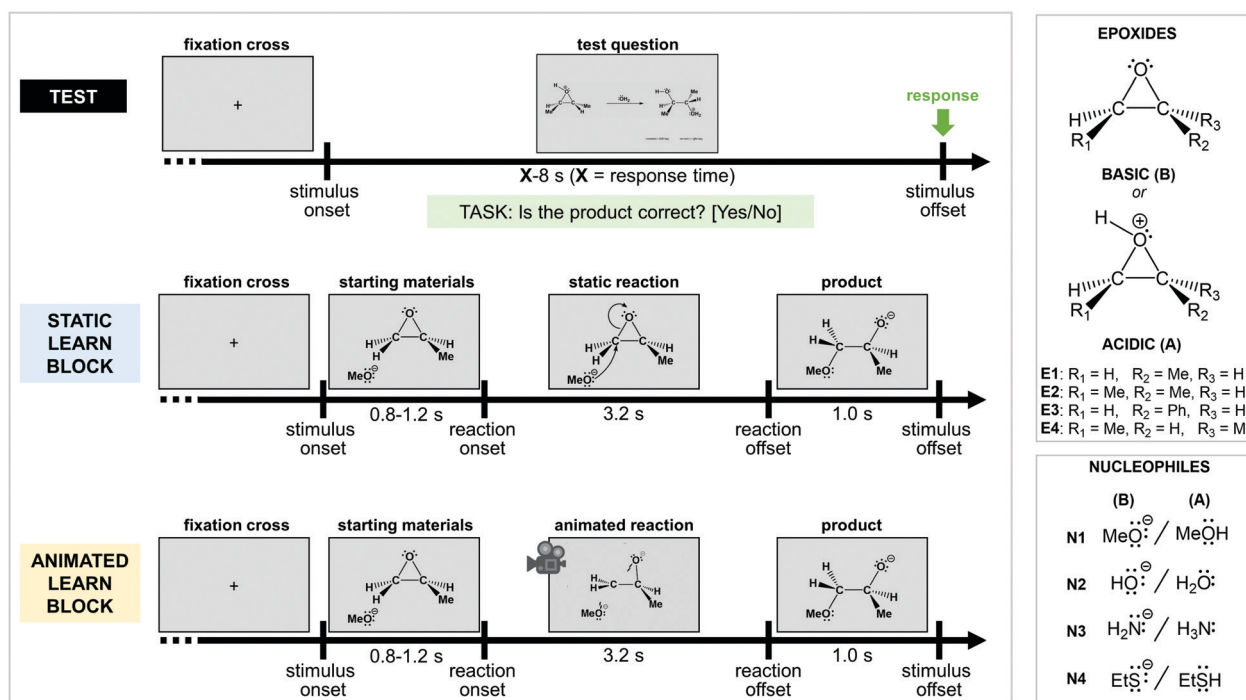
**Table 6** Combinations of epoxides and nucleophiles used in the learn blocks

	Condition	N1	N2	N3	N4
E1	A	X			X <sup>a</sup>
	B	X			X
E2	A		X <sup>a</sup>		
	B		X <sup>a</sup>		
E3	A		X <sup>a</sup>		
	B		X <sup>a</sup>	X	
E4	A	X			
	B				X <sup>a</sup>

<sup>a</sup> This combination was presented twice as mirror images of the reaction.

#### Learn block stimuli

The reaction mechanisms presented in the Learn Block were combinations of four epoxides (E1–E4) and four nucleophiles (N1–N4) under either acidic (A) or basic (B) conditions (Table 6).



**Fig. 11** Experiment details showing one example trial for the Tests and static and animated Learn Blocks. For the animated Learn Block a single screenshot of the animation is shown, and the video camera denotes animation. All trials began with a fixation cross shown for a randomly variable (jittered) time, which served as the inter-trial interval. Epoxides (E1–E4) and nucleophiles (N1–N4) were used in this study to create all stimuli.

**Table 5** Working mental models (WMM) coding scheme

WMM	Description from coding scheme (Bongers <i>et al.</i> , 2019a, 2019b)
Static	Describing their answer in terms of symbols and/or structures, without describing the electron-transfer process, and lacking process-oriented language related to reaction dynamics. Using patterns or heuristics based on symbols ( <i>e.g.</i> , charge signs, substituent positions, and symmetry) and not necessarily on chemical concepts.
Dynamic: process	Describing and/or visualizing the reaction in terms of an electron-transfer process from A to B, in episodic terms ( <i>e.g.</i> , a bond breaks here, bond forms there). Using dynamic step-wise terminology when describing electron transfers.
Dynamic: particles in motion	Describing and/or visualizing the movement of electrons, atoms, or molecules in transition between A and B ( <i>e.g.</i> , a bond is breaking here, bond is forming there). Using dynamic transitional terminology like “flow” or “collide”.



1 **Table 7** Combinations of epoxides and nucleophiles used in the tests

	Condition	N1	N2	N3	N4
E1	A	Regio <sup>a</sup>			Stereo <sup>b</sup>
	B	Regio <sup>a</sup>			
E2	A		Conn <sup>c</sup>		
	B		Conn <sup>c</sup>	Charge <sup>e</sup>	
E3	A		Regio <sup>a</sup>		
	B		Regio <sup>a</sup>	Mech <sup>d</sup>	
E4	A	Stereo <sup>b</sup>			Mech <sup>d</sup>
	B			Stereo <sup>b</sup>	

10 <sup>a</sup> Error: the product had the wrong regiochemistry. <sup>b</sup> Error: the product had the wrong stereochemistry. <sup>c</sup> Error: the product had the wrong connectivity. <sup>d</sup> Error: the product was from the wrong mechanism, either the C–O bond breaks without nucleophilic addition or the nucleophile adds without the C–O bond breaking. <sup>e</sup> Error: the product had the wrong formal charge.

### 15 Test questions

There were 24 test questions representing 12 reaction types. Table 7 shows the combinations of four epoxides (E1–E4) and four nucleophiles (N1–N4) under either acidic (A) or basic (B) conditions used in the test, and which errors were included.

### Abbreviations

EPF	Electron-pushing formalism
LG	Learning gain
PSVT:R	Revised purdue spatial visualization test: rotations
RBR	Rule-based reasoning
CBR	Case-based reasoning
MBR	Model-based reasoning
PiM	Particles in motion
SD	Standard deviation
SE	Standard error
Mdn	Median
<i>M</i>	Mean
MD	Mean difference
<i>B</i>	Unstandardized coefficient beta
$\beta$	Standardized coefficient beta

### Acknowledgements

The authors thank Prof. Ghislain Deslongchamps for creating the animations used in this study and eCampusOntario for funding. Additional thanks to Foster Rose for help with transcribing the think-aloud interviews.

### References

- 50 Akaygun S., (2016), Is the oxygen atom static or dynamic? The effect of generating animations on students' mental models of atomic structure, *Chem. Educ. Res. Pract.*, **17**, 788–807.
- Al-Balushi S. M. and Al-Hajri S. H., (2014), Associating animations with concrete models to enhance students' comprehension of different visual representations in organic chemistry, *Chem. Educ. Res. Pract.*, **15**(1), 47–58.

Aldahmash A. H. and Abraham M. R., (2009), Kinetic versus static visuals for facilitating college students' understanding of organic reaction mechanisms in chemistry, *J. Chem. Educ.*, **86**(12), 1442–1446.

Ardac D. and Akaygun S., (2005), Using Static and Dynamic Visuals to Represent Chemical Change at Molecular Level, *Int. J. Sci. Educ.*, **27**(11), 1269–1298.

Baptista M., Martins I., Conceição T., and Reis P., (2019), Multiple representations in the development of students' cognitive structures about the saponification reaction, *Chem. Educ. Res. Pract.*, **20**, 760–771.

Bhattacharyya G., (2013), From Source to Sink: Mechanistic Reasoning Using the Electron-Pushing Formalism, *J. Chem. Educ.*, **90**(10), 1282–1289.

Bhattacharyya G. and Bodner G. M., (2005), "It Gets Me to the Product": How Students Propose Organic Mechanisms, *J. Chem. Educ.*, **82**(9), 1402–1407.

Bodé N. E., Deng J. M., and Flynn A. B., (2019), Getting Past the Rules and to the WHY: Causal Mechanistic Arguments When Judging the Plausibility of Organic Reaction Mechanisms, *J. Chem. Educ.*, **96**(6), 1068–1082.

Bodner G. M. and McMillen T. L. B., (1986), Cognitive restructuring as an early stage in problem solving, *J. Res. Sci. Teach.*, **23**(8), 727–737.

Bongers A., Flynn A. B., and Northoff G., (2019a), Is learning scale-free? Chemistry learning increases EEG fractal power and changes the power law exponent, *Neurosci. Res. Lett.*, in press, DOI: , DOI: 10.1016/j.neures.2019.10.011.

Bongers A., Northoff G., and Flynn A. B., (2019b), Working with mental models to learn and visualize a new reaction mechanism, *Chem. Educ. Res. Pract.*, **20**(3), 554–569.

Carter C. S., Larussa M. A., and Bodner G. M., (1987), A Study of Two Measures of Spatial Ability as Predictors of Success in Different Levels of General Chemistry, *J. Res. Sci. Teach.*, **24**(7), 645–757.

Caspari I., Kranz D., and Graulich N., (2018), Resolving the complexity of organic chemistry students' reasoning through the lens of a mechanistic framework, *Chem. Educ. Res. Pract.*, **19**(4), 1117–1141.

ChemTube3D: Organic Chemistry Animations, (2018), <https://www.chemtube3d.com/main-page/>, Accessed December 2019.

Christian K. and Talanquer V., (2012), Modes of reasoning in self-initiated study groups in chemistry, *Chem. Educ. Res. Pract.*, **13**(3), 286–295.

Crandell O. M., Kouyoumdjian H., Underwood S. M., and Cooper M. M., (2019), Reasoning about Reactions in Organic Chemistry: Starting It in General Chemistry, *J. Chem. Educ.*, **96**(2), 213–226.

Fanelli D., (2010), Do pressures to publish increase scientists' bias? An empirical support from US states data, *PLoS One*, **5**(4), e10271.

Flynn A. B. and Featherstone R. B., (2017), Language of mechanisms: exam analysis reveals students' strengths, strategies, and errors when using the electron-pushing formalism (curved arrows) in new reactions, *Chem. Educ. Res. Pract.*, **18**(1), 64–77.

- 1 Galloway K. R., Stoyanovich C., and Flynn A. B., (2017), Students' Understanding of Mechanistic Language Prior to Learning Organic Reactions, *Chem. Educ. Res. Pract.*, **18**(2), 353–374.
- 5 Graulich N., Hopf H., and Schreiner P. R., (2010), Heuristic thinking makes a chemist smart, *Chem. Soc. Rev.*, **39**(5), 1503–1512.
- Harle M. and Towns M., (2011), A Review of Spatial Ability Literature, Its Connection to Chemistry, and Implications for Instruction, *J. Chem. Educ.*, **88**(3), 351–360.
- 10 Hegarty M., Stieff M., and Dixon B. L., (2013), Cognitive change in mental models with experience in the domain of organic chemistry, *J. Cogn. Psychol.*, **25**(2), 220–228.
- Höffler T. N., (2010), Spatial Ability: Its Influence on Learning with Visualizations—a Meta-Analytic Review, *Educ. Psychol. Rev.*, **22**(3), 245–269.
- 15 Höffler T. N. and Leutner D., (2007), Instructional animation versus static pictures: a meta-analysis, *Learn. Instr.*, **17**(6), 722–738.
- Höffler T. N. and Leutner D., (2011), The role of spatial ability in learning from instructional animations – Evidence for an ability-as-compensator hypothesis, *Comput. Human Behav.*, **27**(1), 209–216.
- 20 Kelly R. M. and Akaygun S., (2016), Insights into How Students Learn the Difference between a Weak Acid and a Strong Acid from Cartoon Tutorials Employing Visualizations, *J. Chem. Educ.*, **93**(6), 1010–1019.
- 25 Kelly R. M. and Jones L. L., (2007), Exploring How Different Features of Animations of Sodium Chloride Dissolution Affect Students' Explanations, *J. Sci. Educ. Technol.*, **16**(5), 413–429.
- 30 Kelly R. M. and Jones L. L., (2008), Investigating Students' Ability To Transfer Ideas Learned from Molecular Animations of the Dissolution Process, *J. Chem. Educ.*, **85**(2), 303–309.
- 35 Kelly R. M., Akaygun S., Hansen S. J. R. R., and Villalta-Cerdas A., (2017), The effect that comparing molecular animations of varying accuracy has on students' submicroscopic explanations, *Chem. Educ. Res. Pract.*, **18**(4), 582–600.
- 40 Khan Academy: Organic chemistry, (2019), <https://www.khanacademy.org/science/organic-chemistry>, Accessed December 2019.
- Kozma R. B., (2003), The material features of multiple representations and their cognitive and social affordances for science understanding, *Learn. Instr.*, **13**(2), 205–226.
- 45 Kozma R. B. and Russell J., (1997), Multimedia and understanding: expert and novice responses to different representations of chemical phenomena, *J. Res. Sci. Teach.*, **34**(9), 949–968.
- Kraft A., Strickland A. M., and Bhattacharyya G., (2010), Reasonable reasoning: multi-variate problem-solving in organic chemistry, *Chem. Educ. Res. Pract.*, **11**(4), 281–292.
- 50 Lowe R. (1999), Extracting information from an animation during complex visual learning, *Eur. J. Psychol. Educ.*, **14**(2), 225–244.
- 55 Lowe R., (2004), Interrogation of a dynamic visualization during learning, *Learn. Instr.*, **14**, 257–274.
- Lowe R. and Boucheix J.-M., (2008), Learning from Animated Diagrams: How Are Mental Models Built? *Diagrammatic Representation and Inference*, Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 266–281.
- Maeda Y., Yoon Yoon S., Kim-kang G., and Imbrie P. K., (2013), Psychometric Properties of the Revised PSVT:R for Measuring First Year Engineering Students' Spatial Ability, *Int. J. Eng. Educ.*, **29**(3), 763–776.
- Mathôt S., Schreij D., and Theeuwes J., (2012), OpenSesame: an open-source, graphical experiment builder for the social sciences, *Behav. Res. Methods*, **44**(2), 314–324.
- 10 McClary L. and Talanquer V., (2011), Heuristic Reasoning in Chemistry: making decisions about acid strength, *Int. J. Sci. Educ.*, **33**(10), 1433–1454.
- 15 Moreira P., Marzabal A., and Talanquer V., (2019), Using a mechanistic framework to characterise chemistry students' reasoning in written explanations, *Chem. Educ. Res. Pract.*, **20**, 120–131.
- Organic ChemWare, <https://organicchemware.nelson.com/>, Accessed December 2019, (<http://www.nelson.com/organicchemware/>). 20
- Pribyl J. R. and Bodner G. M., (1987), Spatial ability and its role in organic chemistry: a study of four organic courses, *J. Res. Sci. Teach.*, **24**(3), 229–240.
- Ryoo K. and Linn M. C., (2014), Designing guidance for interpreting dynamic visualizations: generating versus reading 25 explanations, *J. Res. Sci. Teach.*, **51**(2), 147–174.
- Scalco K. C., Talanquer V., Kiill K. B., and Cordeiro M. R., (2018), Making Sense of Phenomena from Sequential Images versus Illustrated Text, *J. Chem. Educ.*, **95**(3), 347–354.. **Q5** 30
- Stieff M., (2004), *A localized model of spatial cognition in chemistry*, PhD Dissertation, Northwestern University.
- Stieff M., (2011a), Improving representational competence using molecular simulations embedded in inquiry activities, *J. Res. Sci. Teach.*, **48**(10), 1137–1158. 35
- Stieff M., (2011b), When is a molecule three dimensional? A task-specific role for imagistic reasoning in advanced chemistry, *Sci. Educ.*, **95**(2), 310–336.
- Stieff M. and Rajé S., (2010), Expert algorithmic and imagistic problem solving strategies in advanced chemistry, *Spat. Cogn. Comput.*, **10**(1), 53–81. 40
- Stieff M. and Wilensky U., (2003), Connected Chemistry, Incorporating Interactive Simulations into the Chemistry Classroom, *J. Sci. Educ. Technol.*, **12**(3), 285–302.
- Stieff M., Hegarty M., and Deslongchamps G., (2011), Identifying 45 Representational Competence With Multi-Representational Displays, *Cogn. Instr.*, **29**(1), 123–145.
- Stieff M., Ryu M., Dixon B., and Hegarty M., (2012), The Role of Spatial Ability and Strategy Preference for Spatial Problem Solving in Organic Chemistry, *J. Chem. Educ.*, **89**(7), 854–859. 50
- Stieff M., Dixon B. L., Ryu M., Kumi B. C., and Hegarty M., (2014), Strategy training eliminates sex differences in spatial problem solving in a stem domain, *J. Educ. Psychol.*, **106**(2), 390–402.
- Suits J. P., (2015), Design of Dynamic Visualizations to Enhance Conceptual Understanding in Chemistry Courses, *Chem. Educ. Best Pract. Oppor. Trends*, 595–619. 55

- 1 Suits J. P. and Sanger M. J., (2013), Dynamic visualizations in chemistry courses, in Suits J. P. and Sanger M. J. (ed.), *ACS Symposium Series*, ch. 1, vol. 1142, pp. 1–13.
- 5 Supasorn S., Suits J. P., Jones L. L., and Vibuljan S., (2008), Impact of a pre-laboratory organic-extraction simulation on comprehension and attitudes of undergraduate chemistry students, *Chem. Educ. Res. Pract.*, **9**(2), 169–181.
- 10 Taber K. S., (2013), Revisiting the chemistry triplet: drawing upon the nature of chemical knowledge and the psychology of learning to inform chemistry education, *Chem. Educ. Res. Pract.*, **14**(2), 156–168.
- 15 Tasker R. and Dalton R., (2006), Research into practice: visualisation of the molecular world using animations, *Chem. Educ. Res. Pract.*, **7**(2), 141–159.
- 20 Tasker R. and Dalton R., (2008), Visualizing the Molecular World - Design, Evaluation, and Use of Animations, in Gilbert J. K., Reiner M., and Nakhleh M. (ed.), *Visualization Theory and Practice in Science Education*, Dordrecht: Springer, pp. 103–131.
- 25 Velázquez-Marcano A., Williamson V. M., Ashkenazi G., Tasker R., and Williamson K. C., (2004), The use of video demonstrations and particulate animation in general chemistry, *J. Sci. Educ. Technol.*, **13**(3), 315–323.
- 30 Vlacholia M., Vosniadou S., Roussos P., Salta K., Kazi S., Sigalas M., and Tzougraki C., (2017), Changes in visual/spatial and analytic strategy use in organic chemistry with the development of expertise, *Chem. Educ. Res. Pract.*, **18**(4), 763–773.
- 35 Wang C.-Y. and Barrow L. H., (2011), Characteristics and Levels of Sophistication: An Analysis of Chemistry Students' Ability to Think with Mental Models, *Res. Sci. Educ.*, **41**(4), 561–586.
- 40 Weinrich M. L. and Talanquer V., (2016), Mapping students' modes of reasoning when thinking about chemical reactions used to make a desired product, *Chem. Educ. Res. Pract.*, **17**(2), 394–406.
- 45 Williamson V. M. and Abraham M. R., (1995), The effects of computer animation on the particulate mental models of college chemistry students, *J. Res. Sci. Teach.*, **32**(5), 521–534.
- 50 Wu H.-K. and Shah P., (2004), Exploring visuospatial thinking in chemistry learning, *Sci. Educ.*, **88**(3), 465–492.
- 55 Yang E., Andre T., Greenbowe T. J., and Tibell L., (2003), Spatial ability and the impact of visualization/animation on learning electrochemistry, *Int. J. Sci. Educ.*, **25**(3), 329–349.