Discourse Processes

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/hdsp20

Unpacking the Gestures of Chemistry Learners: What the Hands Tell Us About Correct and Incorrect Conceptions of Stereochemistry

Raedy Ping, R. B. Church, Mary-Anne Decatur, Samuel W. Larson, Elena Zinchenko & Susan Goldin-Meadow

To cite this article: Raedy Ping, R. B. Church, Mary-Anne Decatur, Samuel W. Larson, Elena Zinchenko & Susan Goldin-Meadow (2021) Unpacking the Gestures of Chemistry Learners: What the Hands Tell Us About Correct and Incorrect Conceptions of Stereochemistry, Discourse Processes, 58:3, 213-232, DOI: 10.1080/0163853X.2020.1839343

To link to this article: https://doi.org/10.1080/0163853X.2020.1839343

DISCOURSE PROCESSES
2021, VOL. 58, NO. 3, 213–232
https://doi.org/10.1080/0163853X.2020.1839343

Unpacking the Gestures of Chemistry Learners: What the Hands Tell Us About Correct and Incorrect Conceptions of Stereochemistry

Raedy Ping, R. B. Church, Mary-Anne Decatur, Samuel W. Larson, Elena Zinchenko, and Susan Goldin-Meadow

Department of Psychology University of Chicago; Department of Psychology Northeastern Illinois University

ABSTRACT

In this study, adults naïve to organic chemistry drew stereoisomers of molecules and explained their drawings. From these explanations, we identified nine strategies that participants expressed during those explanations. Five of the nine strategies referred to properties of the molecule that were explanatorily *irrelevant* to solving the problem; the remaining four referred to properties that were explanatorily *relevant* to the solution. For each problem, we tallied which of the nine strategies were expressed within the explanation for that problem and determined whether the strategy was expressed in speech only, gesture only, or in both speech and gesture within the explanation. After these explanations, all participants watched the experimenter deliver a 2-minute training module on stereoisomers. Following the training, participants repeated the drawing + explanation task on six new problems. The number of relevant strategies that participants expressed in speech (alone or with gesture) before training did *not* predict their post-training scores. However, the number of relevant strategies participants expressed *in gesture only* before training *did* predict their post-training scores. Conveying relevant information about stereoisomers uniquely in gesture prior to a brief training is thus a good index of who is most likely to learn from the training. We suggest that gesture reveals explanatorily relevant implicit knowledge that reflects (and perhaps even promotes) acquisition of new understanding.

Introduction: unpacking the gestures of chemistry learners

For undergraduate students, the course called Organic Chemistry is a gatekeeper to postgraduate education in science and healthcare. Given issues of social inequity within professional fields like medicine, it is critical to develop ways to help more students through this gate. Indeed, under-represented minorities experience particular difficulty in Organic Chemistry, largely because individuals with relatively few economic resources (e.g., marginalized and under-represented students) do not have the background needed for the course (Chen, 2013). Organic Chemistry is difficult for many because solving chemistry problems relies heavily on complex spatial and dynamic thinking. Organic molecules are complex three-dimensional forms, intrinsically and extrinsically dynamic, and not visible to the naked eye. Our long-term goal is to create tools to help students develop accurate representations of the complex, three-dimensional, and dynamic processes and phenomena common in science. Providing students with tools of this sort has the potential to give them a helping hand with Organic Chemistry and perhaps can help to level the playing field between those with few and those with many resources.

CONTACT Susan Goldin-Meadow sgm@uchicago.edu Department of Psychology, University of Chicago, 5848 S University Ave, Chicago, IL 60637.

Supplemental data for this article can be accessed on the publisher's website.

Traditionally, teachers have used spoken language, written text, and two-dimensional diagrams to teach chemistry phenomena, but these instructional tools may not be particularly well suited to teaching dynamic three-dimensional processes. We argue that the hand gestures teachers and students often produce when discussing math and science are a good vehicle for capturing dynamic spatial information (Alibali & Nathan, 2012; Alibali et al., 2011; Hostetter & Alibali, 2008), in part because the hands offer potentially useful affordances and because gesture is not constrained by grammar, vocabulary, or the categorical nature of words (e.g., McNeill, 1992). Unlike three-dimensional models, gestures are free of cost, are always available, and can easily be transferred from one learning situation to the next. From experts to novices, people spontaneously and effortlessly gesture when they talk about science (Atit et al., 2014; Crowder & Newman, 1993; Singer et al., 2008). Compared to traditional tools like spoken language, written text, and diagrams, gestures offer distinct advantages for the many students who struggle with Organic Chemistry simply because talk about the spatial and dynamic processes involved in chemistry is likely to be accompanied by iconic gestures that convey three-dimensional, dynamic information. Examining the gestures that students produce when talking about chemistry could give instructors insight into the struggles their students are having with Organic Chemistry. These insights could, in turn, help instructors devise lessons that can address their students' misunderstandings. We see this enterprise as particularly important for those students under-represented in biomedical and Science, Technology, Engineering and Technology careers (e.g., women and students of color) because they have traditionally struggled in math and science classes (Chen, 2013; Rueckert et al., 2017).

We begin by reviewing learning sciences research on gesture in real-world scientific inquiry, and then turn to developmental psychology research on gesture, cognition, and learning. We draw upon both traditions to motivate our hypotheses about the role gesture plays in chemistry learning.

Speech and gesture establish common ground in collaborative real-world science inquiry

Learning sciences researchers have focused on gesture's role in communities where the practice is to represent and communicate about scientific objects and phenomena in situ (e.g., Becvar et al., 2005; Goodwin, 2000, 2007, 2010). Gestures are transcribed from videotapes of scientists and/or students interacting in the lab, field, or classroom and are analyzed in the context of the accompanying dialogue. For example, Trafton et al. (2006) observed meteorologists and neuroscientists at work and asked them periodically to talk out loud about what they were doing. The scientists produced more iconic gestures when they used spatial language than nonspatial language and more iconic gestures when they talked about dynamic processes than static objects. As another example, Becvar et al. (2005) traced the development of shared understanding about a particular protein (thrombin) over weeks. The lab leader produced an idiosyncratic spontaneous gesture for thrombin and other lab members adopted this gesture. Over time, this so-called thrombin hand became a conventionalized form that every lab member used to demonstrate processes involving thrombin. Gesture can be an integral part of coming to a shared understanding on the fly.

But gestures rarely become conventionalized forms and are almost never verbally named ("thrombin hand" is an exception). Instead, they are produced anew by each speaker on each occasion. Radinsky et al. (2012) found that even sixth grade students use one another's spontaneous gestures in science to learn situations to improve their own understanding of tectonic plate movement. Although students did not realize they were depending on one another's gestures, they did so rapidly, responding to others' gestures with their own gestures (and speech). Undergraduate Organic Chemistry students and instructors may also be able to use gesture to come to a shared understanding, making the study of gesture potentially informative about chemistry learning. In this study, we focus on an individual speaker's gestures and ask what the information conveyed uniquely in gesture can tell us about a student's understanding of chemistry.

Gesture–speech mismatches reflect knowledge in transition in child learners

Church and Goldin-Meadow (1986) asked 5- to 8-year-old children to explain how they solved Piaget (1952) conservation problems. On each problem, the child was shown two identical tall, thin glasses and was asked to confirm that the amount of water in the two glasses was the same. While the child watched, the experimenter then poured the water from one of the glasses into a short, wide bowl. Nonconserving children believe that the amount of water changed when it was poured and typically justify this belief as follows: "The glass has more water because it's taller than the dish." The gestural equivalent of the height comparison strategy is a series of flat palm or point gestures that indicate the (higher) level of water in the glass and the (lower) level of water in the bowl.

Children usually express the same strategy in both speech and gesture, as in the height example just given––but not always. Consider a child whose speech conveys height information but whose gesture conveys information about the widths of the containers––a C-shape mirroring the circumference of each container, combined with speech focusing on the heights of the containers. In this case, gesture is adding unique information to the spoken response (here, information about a second dimension, which is not mentioned in speech). Researchers have found that children who produce these so-called gesture–speech mismatches when explaining their solutions to the conservation task are particularly likely to profit from instruction in conservation (Church et al., 2004; Church & Goldin-Meadow, 1986). This phenomenon has been replicated in child learners tackling a variety of science-related tasks (Goldin-Meadow et al., 1993; Perry et al., 1988; Pine et al., 2004).

One reason that gesture–speech mismatches are related to learning may be because gesture indicates nascent understanding of how to solve a particular problem. For example, a child who talks about the height of the containers in speech while gesturing about their width is starting to attend to the width dimension––an important step in understanding that height and width compensate for one another in determining quantity (that even though the water changed in height, it also changed in width). Attending to both the height and width dimensions is one requirement for understanding conservation of liquid quantity after transformation. Gesture–speech mismatch can therefore be used to glean how a child mentally represents a particular problem and to provide a window into thinking not readily afforded by the child's problem solutions or speech strategies for solving the problem.

Do gesture–speech mismatches reflect adult learners' knowledge?

Adults also produce gesture–speech mismatches when talking about concrete and abstract ideas: reasoning about moral dilemmas (Church et al., 1995); describing pictures of landscapes, abstract art, buildings, people, and machines (Morrell-Samuels & Krauss, 1992); narrating cartoon stories (Beattie & Shovelton, 1999; McNeill, 1992; Rauscher et al., 1996); explaining solutions to the Tower of Hanoi puzzle (Beilock & Goldin-Meadow, 2010; Garber & Goldin-Meadow, 2002; Goldin-Meadow & Beilock, 2010); explaining how gears work (Perry & Elder, 1997); and describing solutions to algebra problems involving continuous and discrete change (Alibali et al., 1999). But do adult gestures reveal nascent problem-solving strategies, and do these gestures indicate when an adult is ready to learn how to solve a complex problem?

Adult experts and novices produce iconic gestures during science problem-solving, particularly when working on organic chemistry problems that rely heavily on spatial and dynamic reasoning (Stieff, 2007, 2011; Stieff & Raje, 2010). One such concept is stereochemistry. Stereoisomers are chiral objects[1] not superimposable in real space because they are mirror images of one another across some plane. This geometric notion is often demonstrated by comparing the finger structure of the left and right hands––placing the left hand (palm down) directly on top of the right hand (also palm down), or vice versa, reveals that the two are *non-superimposable* spatial arrangements of the same fingers. Structurally, the left hand is a mirror image of the right hand, and neither hand is symmetric (i.e., all 5 fingers are different and cannot be superimposed). Like the fingers of a hand, the substituents of stereoisomers lack an internal plane or point of symmetry: They contain a central atom called a stereocenter from which extend four different substituent groups in a tetrahedron. Conversely, all parts of *superimposable* molecules can be perfectly lined up with one another, as if the fingers of the left hand were supplanted with those of the right hand. Since molecules are three-dimensional in a way that the hand analogy is not, if any two of the substituent groups in the tetrahedral configuration are identical, the molecule is symmetric and therefore would have a superimposable mate. In other words, it would *not* have a stereoisomer. We chose this task because it is complex and adults have a hard time solving it (but could learn about it relatively quickly) and because it is an excellent representation of the kinds of spatially complex problems students often have to grapple with in an Organic Chemistry course.

From initial field work and discussion with experts, we identified nine commonly used strategies for solving stereoisomer problems. Then, we recruited a sample of adults, naïve to Organic Chemistry, and asked them to try to solve some stereoisomer problems. On each problem, they drew a stereoisomer (if the molecule in the problem had one) and explained their thinking aloud. Later, coders listened to each explanation with a list of the strategies and checked one off as they heard it. On another pass, coders watched each explanation (with no sound) and checked off which strategies were gestured. We then compiled the speech and gesture codes for each explanation and determined whether each problem-solving strategy was produced in that explanation and, if so, whether it was expressed in speech alone, gesture alone, or both speech and gesture.

After solving the first set of problems, participants watched a brief training module about stereochemistry and then solved a second set of problems. We used the explanations each participant produced on the pretraining problems to predict that participant's success on the post-training problems, controlling for performance on the pretraining problems. If adults are like child learners with respect to gesture and learning, adults who produce problem-solving strategies that are expressed only in gesture and not in speech (i.e., gesture–speech mismatches) on the pretraining problems ought to be particularly likely to succeed on the post-training problems.

However, our adult study allows us to take this phenomenon one step further because it was relatively common for the adults to express strategies that were explanatorily irrelevant to solving the problem. In contrast, when children explained how they solved math and conservations problems using incorrect strategies, those incorrect strategies rarely expressed explanatorily irrelevant information. Because the stereoisomer problem is so complex, it has many components that are explanatorily irrelevant to its solution, which a novice might focus on—for example, a longer line drawn between bonds in a molecule does not mean that those bonds are farther away from each other. Even if it did, this property would not be relevant to creating a stereoisomer because it does not change the molecule's spatial arrangement. An analogous response in the child studies might be saying that the cups in a liquid conservation tasks have different amounts of water because the water is dyed blue, a response that is not unheard of but is rare. In contrast, we found that five of the nine strategies produced on the stereoisomer problems were irrelevant to solving the problem; moreover, these irrelevant strategies were produced dozens of times by many participants. We are thus able to ask whether any kind of information emerging uniquely in gesture predicts change (i.e., whether producing any kind of gesture–speech mismatch predicts change) or whether the information uniquely added in gesture in a mismatch must be explanatorily relevant to the problem at hand to predict change. With respect to the stereoisomer problems, we ask whether producing an explanatorily *irrelevant* strategy in gesture and not in speech predicts post-training scores as well as producing an explanatorily *relevant* strategy in gesture and not in speech. Investigating the relationship between modality and relevance will allow us to determine *why* gestured strategies predict post-training scores. Is it the mismatch between the information conveyed in gesture and speech per se that predicts learning, or does the relevance of the information conveyed play a role as well?

To summarize, we provide adult learners with a brief training in stereoisomers and assess their performance on six stereoisomer problems before training and six after training. We describe the explanations the adults give for their problem solutions, focusing on the explanatory relevance of the explanations they express (relevance, irrelevance) and the modality in which each strategy is expressed (speech and gesture, speech only, gesture only). We then use the explanations produced before training to predict scores on the six stereoisomer problems after training, asking three questions: (1) Does strategy relevance, on its own predict, predict post-training scores? (2) Does the modality in which the strategies were expressed, on its own, predict post-training scores? (3) Does the combination of relevance and modality, taken together, predict post-training scores? Our goal is to determine whether the explanatory relevance of a strategy, along with the modality in which it is expressed, predicts the ability to profit from a brief training in adult learners.

Methods

Participants

Fifty-two adults (54% women) participated in the study; 51% of participants were White, 17% Asian American, 15% African American; 15% Latino/a, 6% identified as biracial, 6% responded other, and 4% opted not to report race. All participants were fluent speakers of English and were ages 18 to 22 years. Participants were compensated for their time with either course credit or a small monetary reimbursement. Participants were undergraduates recruited through a list-serve of psychology study volunteers. We prescreened participants for their level of chemistry education and selected volunteers who had at least 1 year of formal chemistry education at the high school or undergraduate level, with no formal instruction in organic chemistry. Five participants (4 women) were excluded from analyses because of experimenter error or noncompliance with instructions.[2] Four additional participants were excluded based on their pretraining scores (3 were unable to give any responses that followed the laws of chemistry and 1 had mastered the task at pretraining).[3] We present data from 43 participants.

Design overview

Each participant followed the same procedure and was given the same problems and training lesson. Every participant answered six questions, then received a brief training, and finally answered six new questions. On each question, participants were given a prompt molecule in a two-dimensional drawing (Figure 1; see also below in *Materials, Prompts of molecules*) and either drew a stereoisomer for the molecule or responded that the molecule does not exist in stereoisomer form. Participants then explained the reasoning behind their drawing (or lack of drawing). We later coded each explanation for the problem-solving strategies it contained and the modality (speech alone, gesture alone, speech and gesture) in which each strategy was produced. We also scored the participants' drawings in response to each problem as "correct" or "incorrect" (see below in *Procedure, Scoring drawings as correct and incorrect*). We use pretraining strategies to predict post-training scores.

Materials

Prompts of molecules

Figure 1 displays an example of a stimulus given to participants in our study. Participants were provided with two different visual representations of molecular structure and spatial arrangement on an 8 × 11 inch piece of paper: a labeled color illustration of a three-dimensional ball and stick representation of the molecule's structure (Figure 1, top) and a wedge and dash representation displaying the molecule's skeletal formula in two dimensions (Figure 1, bottom). Participants were told that the dark-colored triangles (wedges) indicate parts of the molecule coming out of the page in space toward the viewer and that the light-colored triangles (dashes) indicate parts going into the page in space away from the viewer. In this example, the chlorine (Cl) atom is coming out of the page toward the viewer and the lone hydrogen (H) atom is going back into the page away from the viewer. The wedge and dash representation portrays single bonds as single lines between atoms and double bonds as double lines between atoms.[4]

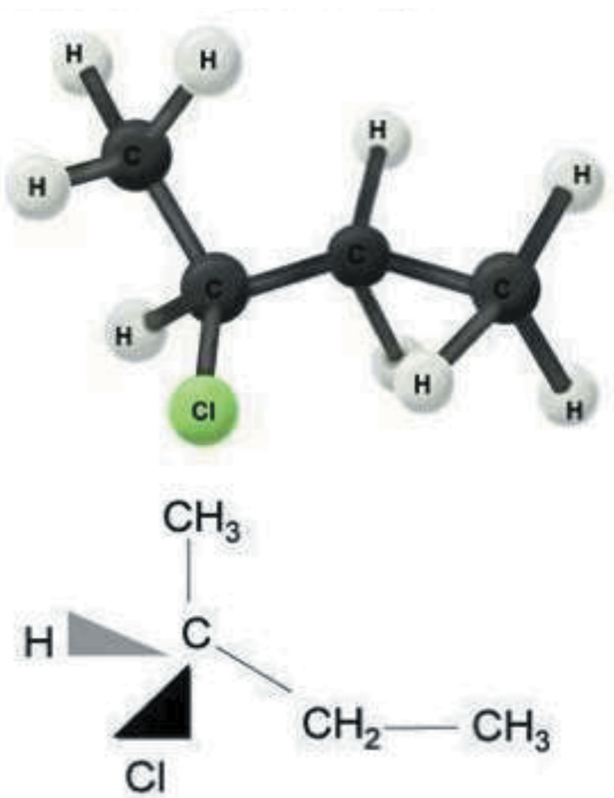


Figure 1. An example of a molecule prompt. *Note*. These images represent the molecule (S)-2-Chlorobutane.

Six unique prompts, one with no stereoisomers of any kind, were presented in a randomized order that was fixed across participants. After the training, a second set of six unique molecules, including one molecule with no stereoisomers, was presented to participants in a randomized order that was fixed across participants. For all participants, the molecule with no stereoisomer was presented as the fourth prompt in the pretraining problems and as the sixth prompt in the post-training problems.

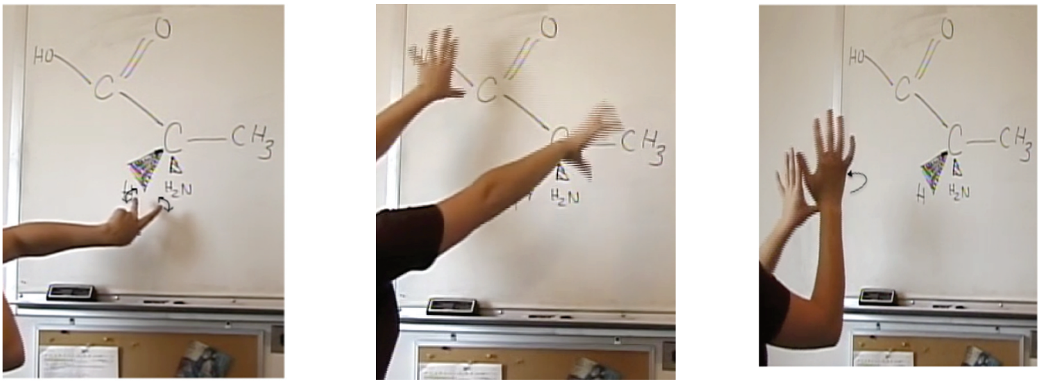
Training

Scripted verbal training was given to participants after they solved the pretraining problems. The experimenter laid out a three-step procedure for determining whether a molecule has a stereoisomer using two example molecules, one with a stereoisomer and one without.[5] The experimenter first indicated the central carbon and the substituents. Participants were then shown how changing the spatial orientation of two of the substituents along the *z*-axis could create a unique spatial representation of the original molecule. Participants were finally told to determine whether the new spatial arrangement actually created a non-superimposable stereoisomer of the original molecule by imagining that they are rotating the entire molecule in three dimensions and checking the nonmanipulated substituents. If the nonmanipulated substituents did not match, it was non-superimposable and therefore a stereoisomer of the original; if they did match, it was superimposable and therefore *not* a stereoisomer.

Procedure

Each individual participated in a single session, lasting less than 1 hour, and followed the same procedure: six problems → training → six problems. Each session was videotaped with the

An Example of Three Relevant Strategies in Speech+Gesture



Relevant Switch	*Compare the Non-Manipulated Substituents*	*Relevant Rotation*
Speech: "Well I *reversed the spatial arrangement of these two*, so now you can't like…	**Speech**: …Since *these are not the same on either side* of the main carbon,	**Speech**: then *you can't like flip it* at all."
Gesture: Index and middle finger each point at one substituent lying at non-zero points on the Z-axis, and wiggle back and forth.	**Gesture**: Left hand flat—points at one substituent on the X-Y plane (where Z=0); right hand flat—points at the other substituent on the X-Y plane.	**Gesture**: Both hands mimic rotation in front of the molecule.

Figure 2. An example of three relevant strategies in speech + gesture.

participants' knowledge, and each participant interacted with the same experimenter throughout the session. When participants arrived, they were given a general description of the problems they would solve and were shown the wedge and dash conventions for drawing molecules. They were then told about the importance of stereoisomers using thalidomide as an example. Stereoisomers were defined for them at a conceptual level as "molecules with multiple non-superimposable spatial arrangements." The experimenter said that stereoisomers are "molecules that have the same molecular formula, as well as the same bond order and connectivity, but different orientations in three-dimensional space." Participants were told "no matter how stereoisomers are rotated in space, their parts are never perfectly superimposable on one another".[6]

Each participant completed 12 total problems: 6 before the training and 6 after. On each problem, the participant was presented with a prompt (Figure 1) and asked to determine whether the molecule exists as a stereoisomer and, if so, to draw a wedge and dash depiction of it on a white board.[7] Participants were then asked to explain how they created their drawing of the stereoisomer. For trials on which participants judged that the molecule does not have a stereoisomer, they were asked to justify why no possible non-superimposable spatial arrangement of the original molecule exists. Participants did not receive feedback on their drawings or their explanations.

Coding and scoring

First, we describe the coding system for the drawings. Next, we describe the process of deriving the coding system for the speech and gesture explanations. Third, we describe how we scored the pretraining and post-training problems using the combination of drawings and speech explanations.

Scoring drawings as correct or incorrect

A correct drawing illustrated a possible stereoisomer of the molecule. For the two prompt molecules without stereoisomers, a drawing was considered correct if the participant stated that the molecule lacked a stereoisomer. A second coder classified 10% of the 516 pretraining and post-training drawings for correctness. Agreement between the two coders was 98%.

Developing speech and gesture coding systems

First, we describe briefly the process of developing the coding systems for the speech and gesture explanations following the drawings: How we identified common strategies and recognized them in each modality. Second, we describe the actual coding we did for the participants in this sample.

Before beginning the study, we videotaped approximately a dozen individuals with varying levels of familiarity with stereochemistry, from psychology graduate students with little to no chemistry education to organic chemistry professors. We asked them to solve different types of stereoisomer problems and to explain their responses in depth. From this corpus, two coders (expert in coding gesture and speech and well versed in stereochemistry) identified common problem-solving strategies. The coders then conducted a series of intensive working meetings with expert gesture coders and chemists to fine-tune the coding system.

Table 1 displays all nine problem-solving strategies identified during this process, along with examples in speech and gesture. These levels reflect increasing understanding of stereoisomers, with level 0 reflecting a lack of mastery and level 4 the highest level of mastery. The top of Table 1 presents the five strategies that are irrelevant to solving the stereoisomer problems; the bottom of Table 1 presents the four strategies that are relevant to the problems.

The first two strategies in Level 0—Changed 2D Representation and No Changes Possible—make incorrect assumptions about chemistry or how molecules work. The next two strategies––Irrelevant Switch and Irrelevant Rotate––focus on the wrong parts of the molecule and therefore are not useful in finding a stereoisomer. Ignoring the Z dimension is also a Level 1 strategy as it omits an important dimension.

The next two strategies––Relevant Switch and Relevant Rotate––are Level 2 and highlight movements performed on appropriate parts of the molecule to determine whether it has a stereoisomer. The Mirror Image strategy is Level 3 because it demonstrates an understanding of the spatial arrangement of the molecule as a whole. But this strategy is not failsafe––if the molecule has internal symmetry, drawing its mirror image is actually just drawing the molecule from a different vantage point. Only Comparing Non-manipulated Substituents, a Level 4 strategy, ensures that the participant fully understands how to solve the problem.

Speech was classified as uncodable on problems where participants did not provide meaningful strategy information in their spoken explanation (e.g., “I am not really sure how I got this but I think it’s right,” “I just did what I did last time”) and, more commonly, on problems where participants repeated the definition of a stereoisomer (e.g., “my drawing is not superimposable on the original molecule”). The experimenter gave a “why?” prompt in response to both kinds of answers; responses were classified as uncodable if participants did not respond to the prompts. Spoken responses were coded with the video turned off with one exception: when it was necessary to view the video in order to determine a referent (e.g., “I moved *this part*” or “I switched *these two groups*”). Speech was codable in 90% of pretraining problems (231/258 problems) and 94% of post-training problems (243/258 problems).

Table 1. Problem-Solving Strategies, Descriptions, and Examples from Speech and Gesture

Strategy Name	Description	Speech Example	Gesture Example
Irrelevant strategies			
Changed two-dimensional representation	By changing the length of the bonds between atoms or altering the angle between two bonds, they have created a stereoisomer.	"I made the bonds between the carbon and the hydrogen shorter."	Pinch two fingers together over the C and H.
No changes possible	Regardless of how they change the molecule it can always be rotated back to the original form and thus cannot create a stereoisomer. Or, due to the laws of chemistry no changes are allowed to be made.	"Anyway I drew it, it would still be super-imposable on the original molecule"	Point or sweep to every substituent and produce no other gestures.
Ignore Z dimension	Indicate having changed the location of two substituents by altering their spatial relation on the two-dimensional X-Y plane of the drawing in such a way that represents no three-dimensional changes.	"Originally the hydrogen was to the bottom left of the carbon, now it's to the bottom right. And the bromine was on the right, now it's on the left."	V-hand shape, pointing to each of the two substituents on the *z*-axis, typically wiggling back and forth.
Irrelevant switch	Participants indicate that rotating two or more substituents that are NOT attached to the stereocenter will produce a stereoisomer.	"I rotated these two CH_3s all around this carbon so now it is different."	Point to a substituent that is not attached to stereocenter, and then sweeps away from the board.
Irrelevant rotate	Participants indicate that rotating two or more substituents that are NOT attached to the stereocenter will produce a stereoisomer.	"I rotated these two CH_3s all around this carbon so now it is different."	Lift up one hand with fingers pointed upward and twist the hand either clockwise or counter-clockwise. The movement is produced in front of a substituent not attached to a stereocenter.
Relevant strategies			
Relevant switch	Participants indicate that no matter how their drawn molecule is rotated, it cannot be superimposed on the original molecule. For items that do not have a stereoisomer, rotating the drawn molecule will return the arrangement to its original form and thus it can be superimposed.	"Because regardless of how you rotate the molecule, the parts of this one are not going to line up with the parts of the original one."	Point with a V-shaped hand to two substituents attached to a stereocenter and flip hand.
Relevant rotate	Participants indicate that no matter how their drawn molecule is rotated, it cannot be superimposed on the original molecule. For items that do not have a stereoisomer, rotating the drawn molecule will return the arrangement to its original form and thus it can be superimposed.	"Because regardless of how you rotate the molecule, the parts of this one are not going to line up with the parts of the original one."	Lift up one hand with fingers pointed upward (or both hands) and twist the hand(s) either clockwise or counter-clockwise. The movement is produced in front of the entire molecule or stereocenter.
Mirror image	Participants indicate that by creating the original molecule's mirror image they have created a stereoisomer.	"If you reflect the molecule over a mirror plane then you couldn't match it up with the original."	Place one hand over one drawing and then flip hand from palm down to palm up over the top of the original molecule.

(*Continued*)

Table 1. (Continued).

Strategy Name	Description	Speech Example	Gesture Example
Check nonmanipulated substituents	Indicate that a molecule manipulated by either mirror image (Level 3) or relevant switch and relevant rotation (Level 2) must be compared to the original to check superimposability. Typically participants checked manipulated substituents against their original orientations, then checked the two nonmanipulated substituents' locations to their original orientations	"When you change which one of these two groups are going into the board and which one is coming out of the board, no matter how you rotate the molecule, it won't be the same. So I rotate this around, these will be back in their original spots, but the two things I didn't change are now in different spots."	Point to one or both of the unchanged substituents attached to the stereocenter.

Gesture was uncodable when participants produced (1) only *beat* gestures (rhythmic, baton-like, up and down movements of the hand and arm that synchronize with speech but have no representational meaning; McNeill, 1992), (2) gestures that were too far away from the drawing to determine the appropriate strategy, or (3) a series of individual disparate points. Gesture was codable in 69% of the pretraining problems on which participants gestured (171/248 problems) and 72% of the post-training problems on which participants gestured (179/246 problems).

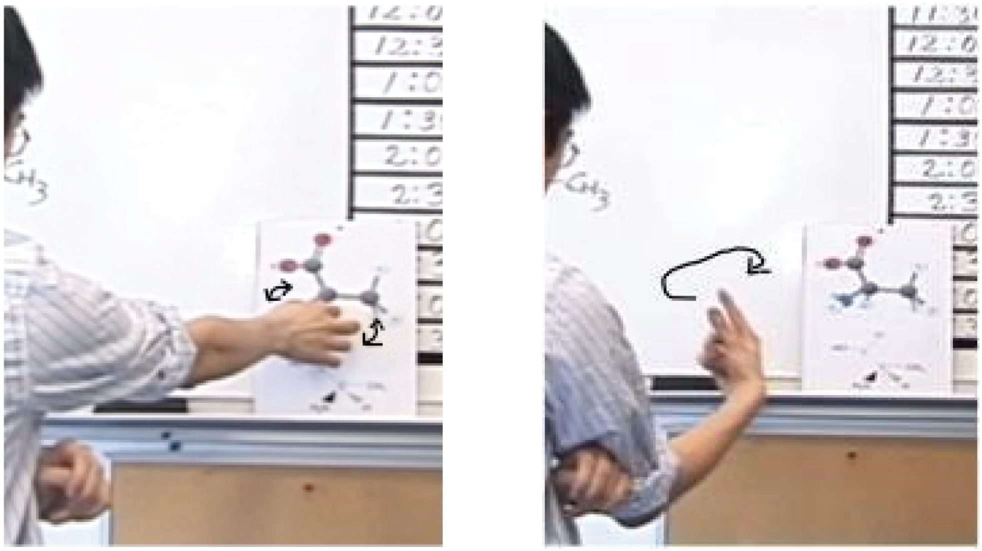
Coding speech and gesture for each stereoisomer problem

One main coder took two separate passes through each explanation. The participant's entire response following each problem was counted as a single explanation, and the speech and gesture contained in that explanation were each classified into one or more of the nine strategies (Table 1; an explanation could contain more than one of the nine strategies). On the first pass, the coder listened to the explanations without looking at the video. She had a list of problem-solving strategies and checked off any strategy or strategies that she heard in the explanation. On a second pass, the coder watched the video without hearing the speech and checked off any strategy or strategies she saw in the gestures. Reliability was assessed by having a second coder go through a randomly selected 10% of our 516 explanations; 10% is typically considered a reasonable sample for inter-rater reliability in qualitative research where complex responses are coded (see Campbell et al., 2013; O'Connor & Joffe, 2020), particularly since we used 10% of explanations, not participants. Agreement between the two coders was 91% for speech strategies and 82% for gesture strategies.

Take, for example, an explanation in which the strategies were expressed in both speech and gesture. In the speech pass, the coder hears "Well, I *reversed the spatial arrangement of these two*, so now you can't like ... Since *these are not the same on either side of the main carbon*, then you *can't like flip it* at all." The three strategies identified in the speech are italicized here for exposition. On the spoken strategy checklist, the coder marks three strategies: "relevant switch," "compare nonmanipulated substituents," and "relevant rotate." On a different day, the coder watches the video with no sound and no access to the speech coding. The participant *puts up two fingers at two relevant substituents and waves the fingers*, then *points at the far ends of the molecule*; finally, *moves her hand as though spinning a relatively large object about a vertical axis*. On the gesture strategy checklist, the coder marks three strategies "relevant switch," "compare nonmanipulated substituents," and "relevant rotate." The final step is to compare the speech and gesture checklists. Three strategies were expressed in this explanation, and all three were expressed in both speech and gesture (see Figure 2).

As an example of an explanation in which one strategy was expressed in both speech and gesture and one was expressed uniquely in gesture, consider the participant in Figure 3. In the speech pass, the coder hears "There's an NH_2 group and an H group connected to this carbon atom. *If you switch these two groups*

An Example of a Relevant Strategy in Speech+Gesture (Left) and a Relevant Strategy in Gesture Alone (Right)



Relevant Switch (in gesture + speech)	*Relevant Rotation* (in gesture alone)
Speech: "There's an NH_2 group and an H group *connected to this carbon atom. If you switch these two groups around* then	**Speech**: according to my head, you can't superimpose this on top of that."
Gesture: Index and middle finger each point at one substituent lying at non-zero points on the Z-axis, and wiggle back and forth.	**Gesture**: Finger pointed upward and hand rotates counter clockwise twice. The movement is produced in front of the entire molecule generally (and not a specific substituent)

Figure 3. An example of a relevant strategy in speech + gesture (left) and a relevant strategy in gesture alone (right).

around, then, according to my head, you cannot superimpose this on top of that." The coder marks one strategy "relevant switch" on the speech checklist. In the gesture pass, the coder sees the participant *wiggle two fingers back and forth* over two relevant substituents and then point his finger toward the ceiling and *rotate his wrist*. The coder marks two strategies on the gesture checklist: "relevant switch" and "relevant rotate." Comparing the two checklists, we find that two strategies were expressed in this explanation––one in speech and gesture ("relevant switch"), and one in gesture only ("relevant rotate").

Scoring each problem as correct or incorrect

We scored a problem as correct when it had both a correct drawing and a Level 4 explanation in speech. Producing a Level 4 explanation indicates that the participant understands how all the constituents in the molecule are spatially related to one another, regardless of how the molecule moves in space. When produced with a correct drawing, the response indicates that the participant has a complete understanding of the stereoisomer.

To be certain that our findings were not determined by our criterion for correctness, we also scored all problems using a looser criterion. Under this scoring, a problem was considered correct if it had a correct drawing and a Level 3 or higher explanation in speech. In order to produce a Level 3 explanation, the participant must know that reflecting the entire molecule across an axis is one way to create a chiral isomer, which shows some understanding of stereoisomers. However, when the molecule is symmetric along one axis, mirroring results in the same molecule, it is essential to check all of the substituents (Level 4) to ensure that the molecule has a stereoisomer. The maximum number of pretraining problems that a participant could solve correctly was six on the pretraining assessment and six on the post-training assessment.

Review of variables

Pretraining and post-training problems

Success was scored with a strict criterion (correct drawing + Level 4 explanation in speech; data are presented in the text) and a loose criterion (correct drawing + Level 3 or 4 explanation in speech; data are presented in supplementary materials and Figure 4, bottom).

Strategies produced to explain the pretraining and post-training drawings

Participants produced nine different types of strategies to explain their drawings, five that were explanatorily irrelevant to the problem and four that were explanatorily relevant (Table 1). An explanation could contain more than one strategy. Each strategy in an explanation was coded for explanatory relevance (irrelevant, relevant) and modality (expressed in both gesture and speech, expressed only in speech, expressed only in gesture), resulting in six distinct coding categories.

We assessed the total number of strategies (summed across participants), categorized according to relevance and modality, that were produced before and after training (Table 2), as well as the change in mean number of each type of strategy from pre- to post-training (Table 3). We also used the strategies on the pretraining problems to predict scores on the post-training problems. We first used the relevance of the strategies (collapsed over modality) and the modality of the strategies (collapsed over relevance) as predictors of post-training problems (Table 4). We then combined the factors and used relevance *and* modality as joint predictors of post-training problems (Table 5, Figure 4).

Results

Strategies expressed on the pretraining and post-training problems

All participants gestured when explaining at least one problem. They produced gesture on 96% of pretraining problems (248/258) and 94% of post-training problems (246/258). They produced speech on all pretraining and post-training problems. Participants often produced more than one strategy within a single explanation: 42 of 43 participants produced at least one explanation on the pretraining problems that included more than one strategy (either in the same modality or in two different modalities), producing an average of 1.71 strategies, $SD = 0.54$, per explanation.

Total number of strategies, classified according to relevance and modality, expressed before and after training

Table 2 displays the total number of times each of the nine strategies was expressed and the modality or modalities in which it was expressed on the pretraining problems and on the post-training problems.

Mean number of strategies, classified according to relevance and modality, expressed before and after training

Table 3 compares strategies produced prior to training to strategies produced after training. The top of the table presents the mean number of irrelevant strategies produced per participant across the six

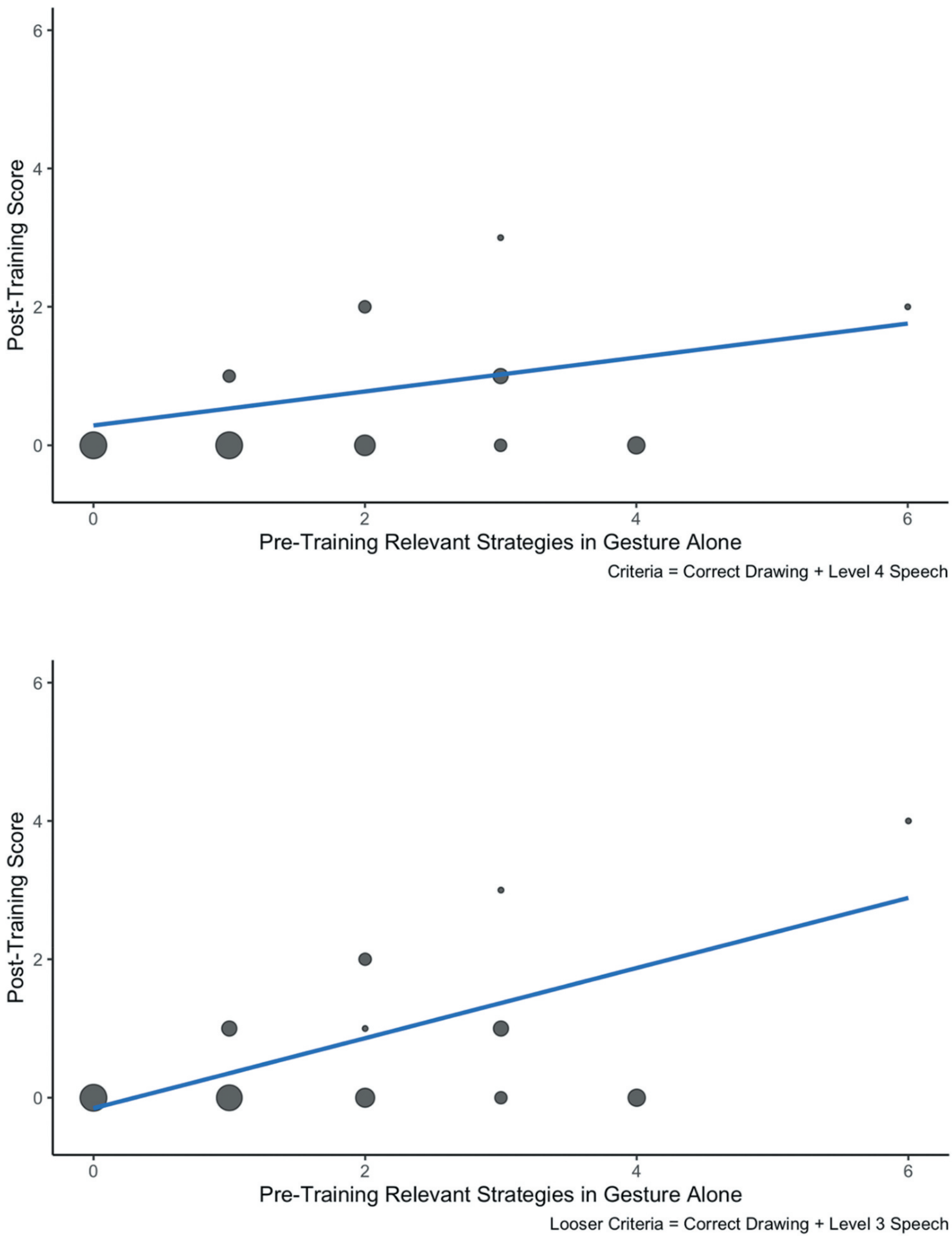


Figure 4. Scatter plot of post-training score as a function of pretraining relevant strategies in gesture alone with stricter criteria (top) and looser criteria (bottom). *Note*. Post-training module score is determined by the number of correct drawings + Level 4 explanations in speech in the top graph and by the number of correct drawings + Level 3 or 4 explanations in the bottom graph. The size of the dot in each graph indicates the number of participants represented by that dot.

problems, classified according to the modality or modalities in which each strategy was produced, and the bottom of the table presents the mean number of relevant strategies. If the training had an effect on participants' understanding of stereoisomers, we would expect irrelevant strategies to decrease and relevant strategies to increase after training. In fact, we found that irrelevant strategies decreased overall from pretraining to post-training; the decrease was significant for irrelevant strategies produced in Speech+Gesture, $t(42) = -2.78$, $p = .008$. We also found that relevant strategies increased

Table 2. Number of Instances of Each Strategy, Summed Across All Participants and All Problems

Strategy Name	Pretraining Problems: Modality: In Speech + Gesture	In Speech Alone	In Gesture Alone	Sum, Any Modality	Post-Training Problems: Modality: In Speech + Gesture	In Speech Alone	In Gesture Alone	Sum, Any Modality
Irrelevant strategies								
Sum, Irrelevant Strategies	*45*	*103*	*50*	*198*	*22*	*68*	*35*	*125*
Changed two-dimensional representation	1	2	14	*17*	0	2	7	*9*
No changes possible	0	2	13	*15*	1	11	1	*13*
Ignoring Z dimension	15	36	5	*56*	9	21	4	*34*
Irrelevant switch	21	54	6	*81*	10	28	4	*42*
Irrelevant rotate	8	9	12	*29*	2	6	19	*27*
Relevant strategies								
Sum, relevant strategies	*101*	*74*	*69*	*244*	*158*	*163*	*50*	*371*
Relevant switch	57	29	30	*116*	104	60	32	*196*
Relevant rotate	41	23	37	*101*	49	83	15	*147*
Mirror image	2	18	0	*20*	0	5	0	*5*
Nonmanipulated substituents	1	4	2	*7*	5	15	3	*23*

Sums are represented as italicized numbers.

Table 3. Number of Instances of Each Strategy, Summed Across Problems and Averaged Across Participants

Strategy Relevance	Strategy Modality: In Speech + Gesture: Pretraining	Post-Training	*t*	*p*	In Speech Alone: Pretraining	Post-Training	*t*	*p*	In Gesture Alone: Pretraining	Post-Training	*t*	*p*
Irrelevant strategies	1.05 (1.37)	0.51 (0.86)	−2.78	.008*	2.65 (2.41)	1.58 (2.41)	−2.55	.014	0.91 (1.29)	0.81 (1.24)	−0.35	.73
Relevant strategies	2.35 (1.97)	3.67 (2.45)	3.38	.001*	1.72 (1.54)	3.79 (2.41)	5.69	<.0001*	1.60 (1.45)	1.16 (1.13)	−1.63	.11

* Family-wide *p* value set to alpha = .05.

from pretraining to post-training; the increase was significant for relevant strategies produced in Speech+Gesture, $t(42) = 3.38$, $p < .001$, and in Speech Only, $t(42) = 5.69$, $p < .0001$.

Using pretraining strategies to predict scores on post-training problems

Does pretraining strategy relevance, on its own, predict scores on post-training problems?

We next explored whether the relevance of strategies (relevant or irrelevant) produced on the pretraining problems predicted success on the post-training problems. Table 4 presents the model

Table 4. Model Summary for Binary Mixed Models Predicting Score on Post-Training Problems from Strategy Relevance (Top) or Strategy Modality (Bottom)

	β	*SE*	*t*	*p*
Model for strategy relevance				
(Intercept)	−4.8439	1.2503	−3.874	.0001
Score on pretraining problems	6.5227	5.2862	1.238	.2157
Relevant strategies, any modality	0.9205	0.8334	1.104	.2694
Irrelevant strategies, any modality	0.2913	0.7370	0.395	.6926
Model for strategy modality				
(Intercept)	−4.1428	1.6722	−2.477	.0132
Score on pretraining problems	11.9660	8.066	1.483	.1380
Strategies in speech + gesture	−0.5746	1.3234	−0.434	.6641
Strategies in speech alone	−0.8332	1.4702	−0.567	.5709
Strategies in gesture alone	2.0072	1.2264	1.637	.1017

Table 5. Model Summary for Binary Mixed Model Predicting Score on Post-Training Problems from Strategy Relevance and Strategy Modality

(Intercept)	−2.056	1.817	−1.131	0.2580
Score on Pre-Training Problems	18.688	8.624	2.167	0.0302*
Relevant Strategies in Speech+Gesture	−1.812	1.741	−1.041	0.2981
Relevant Strategies in Speech Alone	−3.701	2.752	−1.345	0.1786
Relevant Strategies in Gesture Alone	4.385	1.698	2.583	0.0098*
Irrelevant Strategies in Speech+Gesture	3.565	2.522	1.414	0.1575
Irrelevant Strategies in Speech Alone	−4.623	3.000	−1.541	0.1234
Irrelevant Strategies in Gesture Alone	−2.626	2.054	−1.279	0.2010

* Significant predictor at the $p < .05$ level.

summaries for two separate models. The top of Table 4 presents a model predicting the binomial likelihood of getting a post-training problem correct (as measured by producing a correct drawing and a Level 4 explanation in speech). The model includes a random effect of participant; fixed effects were participant's pretraining score, number of pretraining relevant strategies (collapsed over modality), and number of pretraining irrelevant strategies (collapsed over modality). Neither the number of irrelevant, p =.69 (see top of Table 4), nor the number of relevant strategies, $p = .27$ (see top of Table 4), produced prior to training significantly predicted post-training score.

Does strategy modality, on its own, predict scores on post-training problems?

We then asked whether the modality in which a strategy was expressed (speech + gesture, speech alone, gesture alone) produced on the pretraining problems predicted success on the post-training problems. The bottom of Table 4 presents a model predicting the binomial likelihood of getting a post-training problem correct (as measured by producing a correct drawing and a Level 4 explanation in speech). The model includes a random effect of participant; fixed effects were participant's pretraining score, number of strategies expressed in speech + gesture (collapsed over strategy type), number of strategies expressed in speech alone (collapsed over strategy type), and number of strategies expressed in gesture alone (collapse over strategy type). The number of strategies expressed in speech + gesture, $p = .66$ (see bottom of Table 4), speech alone, $p = .57$ (see bottom of Table 4), or gesture alone, $p = .10$ (see bottom of Table 4) were not significant predictors of post-training score.

Loosening the criterion for correctness. To ensure that our findings were not dependent on the criterion for correctness that we used, we redid the analyses in Table 4 using a looser criterion for correctness (a correct drawing + a Level 3 or 4 explanation in speech). Using this criterion, we found that all the results in Table 4 remained the same, except that the number of gesture-only strategies (ignoring relevance) on the pretraining problems predicted success on the post-training problems, $p = .04$.

Does the combination of relevance and modality predict scores on post-training problems?

Our final step was to combine relevance of strategy (relevant, irrelevant) and modality (speech + gesture, speech only, gesture only) and ask whether strategies produced on the pretraining problems, categorized according to type of strategy *and* relevance predicted success on the post-training problems. Table 5 presents a model predicting the binomial likelihood of getting a post-training problem correct (as measured by producing a correct drawing and a Level 4 explanation in speech). The model includes a random effect of participant; fixed effects were participant's pretraining score, number of relevant strategies expressed in speech + gesture, number of relevant strategies expressed in speech alone, number of relevant strategies expressed in gesture alone, number of irrelevant strategies expressed in speech + gesture, number of irrelevant strategies expressed in speech alone, and number of irrelevant strategies expressed in gesture alone. We found that the pretraining score was a significant predictor of the post-training score. The only other significant predictor of the post-training score was the number of *relevant* strategies expressed in *gesture only,* $p = .009$ (see Table 5). Figure 4 (top) displays a scatter plot of post-training scores (y-axis) as a function of the number of relevant strategies expressed in gesture only prior to training (x-axis); the size of each dot represents the number of participants at that point. The scatterplot shows the positive correlation between the two variables.

Loosening the criterion for correctness. We redid the analyses using the looser criterion for correctness (a correct drawing + a Level 3 or 4 explanation in speech) and found the same results––number of relevant strategies expressed in gesture only was a significant predictor of post-training score, as was pretraining score (see Table S1 in supplementary materials). Figure 4, bottom, displays the scatter plot with success on the post-training problems (y-axis) calculated in terms of the looser criterion (correct drawing = Level 3 or 4 explanation in speech); again, the size of each dot represents the number of participants at that point. Even using the looser criterion, we see a positive correlation between relevant strategies produced only in gesture and post-training score.

Discussion

Our study has four central findings. First, we demonstrated that novices gestured at a high rate when they talked about how they transformed molecules in an attempt to create alternate spatial arrangements of organic compounds. Our findings thus replicate studies showing that tasks involving mental rotation and other visuospatial skills tend to elicit gesture (Chu & Kita, 2011; Ehrlich et al., 2006; Stieff, 2007, 2011; Stieff & Raje, 2010; Trafton et al., 2006). Second, we developed a coding system for identifying problem-solving strategies in speech and the accompanying spontaneous gestures. We believe this paradigm and coding system offer promise for studying the role of spontaneous gesture in chemistry learning. Third, we found that the strategies expressed in gesture did not always overlap with those expressed in speech. Fourth, we found that strategies expressed in gesture only prior to training significantly predicted success on the stereoisomer problems after training.

Previous work with children has shown that learners on the brink of conceptual change often produce information uniquely in gesture—information not found in the accompanying speech, so called gesture–speech mismatches (Alibali & Goldin-Meadow, 1993; Church & Goldin-Meadow, 1986; Perry et al., 1988; Pine et al., 2004; for review, see Goldin-Meadow, 2003). In previous work, children were asked to learn relatively straightforward tasks (e.g., Piagetian conservation tasks in 5- to 8-year-olds, mathematical equivalence tasks in 9- to 10-year-olds). The incorrect information that the children conveyed uniquely in gesture on these tasks always contained information that was relevant to the task or concept to be mastered. For example, on the mathematical equivalence problem 4 + 6 + 3 = __ + 3 used by Perry et al. (1988), the correct answer is 10. However, children often respond with "13" to this problem and, when asked them how they got this answer said, "I added the 4, the 6, and the 3" (*add-to-equal-sign* strategy). While producing this strategy in speech, many children produce gestures that convey a different incorrect strategy––they point at the 4, the 6, the left 3, and then the right 3 (*add-all-numbers* strategy), which gives 16 instead of 13. Although the strategies are not correct, both are *relevant* to solving the problem. To solve the problem correctly, children need to notice that the equals sign breaks the problem into two sides (*add-to-equal sign*) and also notice that there is an extra addend on the right side of the equation (*add-all-numbers*). Children who produce these two strategies, one in gesture and the other in speech, are particularly likely to learn how to solve the math problems after a brief training.

Producing information in gesture that is not in speech could index later understanding for at least two reasons. The first is that having several strategies activated at the same time, with some strategies expressed in speech and others in gesture, reflects a type of general cognitive instability in the learner. During such times of instability, learners may be more amenable to instruction in new strategies, possibly because they have become aware of inconsistencies in their thinking and are working to resolve those inconsistencies (Goldin-Meadow & Alibali, 2002; Goldin-Meadow et al., 1993). Variability in strategy choice, either within or across modalities and regardless of whether the strategies are correct or incorrect, has been shown to precede learning (for review see Siegler, 2007;

but see Church, 1999). Under this hypothesis, what is important in propelling change is that multiple strategies be simultaneously activated, which puts the learner in an unstable state. In other words, the relevance of the strategies should not matter.

The second possibility is that gestured information represents implicit knowledge that the learner holds but is, as yet, unable to explicitly express in speech (Broaders et al., 2007). Under this possibility, the type of information embodied in the gestural component of a mismatch ought to be an important index of whether the learner is open to change. When explanatorily irrelevant strategies are expressed in gesture only, gesture reflects implicit ideas that the learner holds, but those ideas are hindering the ability to improve on the task—or at least they are not helping. In this account, conveying information in gesture that is not found in speech is a good index of openness to instruction *only if* the information conveyed in gesture is explanatorily relevant, and a step on the road, to solving the problem correctly.

Our data align with this second account, providing support for the view that the type of information gesture adds to speech in a mismatch is telling. Gesture–speech mismatch likely is a signal of general cognitive instability and thus readiness for change. However, our results suggest that the explanatory relevance of the information conveyed in gesture indicates whether change for the better is likely to come easily to a given individual. In other words, mismatch on its own is not enough––the mismatches learners produce must convey information that is *explanatorily relevant* to solving the problem in order for them to profit from input and change for the better.

Of course, it is possible that findings from this work apply only to problems involving spatial arrangements of molecules, or to problems of molecule transformation, or even to the specific type of enantiomer task we use in this paradigm. Solving the task in our study relies on spatial processes––the participants produced iconic gesture at very high rates, replicating findings from the chemistry education literature showing that novices (and experts, when problems are difficult) approach problems of this sort spatially (Stieff, 2011; Stieff & Raje, 2010). There is a strong possibility that the effects of gesture on learning are more pronounced for problems that have a spatial, or specifically rotation, component (Chu & Kita, 2011; Cook & Tanenhaus, 2009; Hostetter & Alibali, 2008). However, our data also are consistent with a general gesture–speech mismatch phenomenon, established in the literature across a variety of content areas, including moral reasoning problems with no inherent spatial properties (Beaudoin-Ryan & Goldin-Meadow, 2014). The regularity of this effect across studies helps to strengthen the argument that gesture is an important and additional lens through which to study learning more generally.

Our study also underscores the role that gesture can play in the classroom, offering another window into a student's understanding of stereoisomer problems. Importantly, extensive training is not necessary for teachers to understand their students' gestures. Ordinary listeners are able to decode the information that speakers convey uniquely in gesture (Goldin-Meadow & Sandhofer, 1999), often without conscious awareness of gesture as the source of the information (McNeill et al., 1994). Listeners decode iconic information in gesture online, as it occurs, and incorporate the information into their model of the spoken message, as evidenced by reaction time (Ping et al., 2014). Teachers and undergraduate research participants are sensitive to gesture when identifying which problem-solving strategies are in children's explanations of their answers to math and conservation problems after a brief primer on coding categories (Kelly et al., 2002) and even without the primer (Alibali et al., 1997; Goldin-Meadow et al., 1992). Students' spontaneous gestures are thus accessible to teachers even if the teacher has not been trained in gesture coding.

In sum, we have found that gesture can reveal an individual's understanding of stereoisomer problems––the gestures that people produce when they explain their responses to a task reflect the knowledge they have about the task and, importantly, reveal knowledge that often goes beyond the knowledge they convey in their speech. We have also found that the gestures adults spontaneously produce on a stereoisomer task are a good index of how well they will perform on the task after new input, thus replicating previous work with children and demonstrating that when considered in relation to speech, gesture is a good index of openness to change in all learners, young and old. However, the mismatch between gesture and speech is only one component necessary to predict learning––the information conveyed in the mismatching gesture must also be explanatorily relevant to solving the problem. Overall, our findings make it clear that the gestures chemistry novices spontaneously produce can be used by researchers and teachers alike to gain insight into what learners know about chemistry, and how ready they are to learn more.

Notes

1. Not all stereoisomers are chiral, but in our study we used chiral and chiral-looking molecules.
2. For example, the experimenter forgot to turn on the video camera. In another case, the experimenter did not stop a participant from holding the marker during the explanations, which inhibited gesture.
3. Some participants were not able to produce drawings with the same molecular formula and bonding order as given in the stimulus molecules.
4. There are no double bonds in the molecule in the figure, but some molecules used in the study did contain double bonds.
5. The molecule with a stereoisomer had an enantiomer, a diastereomer, and an enantiomer of the diastereomer.
6. Although there are different types of stereoisomers, in the current study we limited our prompts and instruction materials to molecules with mirror-image configurational enantiomers and, as a contrast, molecules with symmetry that do not exist in enantiomer form. We use the terminology "stereoisomer" and "stereochemistry" throughout. Some of the molecules used as stimuli exist in diastereomer form as well. The same strategies can be used to create a stereoisomer for diastereomers.
7. During each trial, participants were allowed to revise their drawing as many times as they desired but were then asked to restart their explanation from the beginning each time they did. Only the final drawing and explanation were included.

Acknowledgments

We are grateful to Mike Stieff for very helpful comments on a rough draft of the manuscript. We also thank Melissa Herrett and Theodora Kmoutsakis for help with figures.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This work was supported by NICHD R01-HD47450 and NSF BCS-0925595 to SGM and NSF SBE-0541957 (Spatial Intelligence and Learning Center) to Nora Newcombe (PI) and SGM (co-PI).

References

Alibali, M. W., Bassok, M., Olseth, K. L., Syc, S. E., & Goldin-Meadow, S. (1999). Illuminating mental representations through speech and gesture. *Psychological Sciences*, *10*(4), 327–333. https://doi.org/10.1111/1467-9280.00163

Alibali, M. W., Flevares, L. M., & Goldin-Meadow, S. (1997). Assessing knowledge conveyed in gesture: Do teachers have the upper hand? *Journal of Educational Psychology*, *89*(1), 183–193. https://doi.org/10.1037/0022-0663.89.1.183

Alibali, M. W., & Goldin-Meadow, S. (1993). Gesture-speech mismatch and mechanisms of learning: What the hands reveal about a child's state of mind. *Cognitive Psychology*, *25*(4), 468–523. https://doi.org/10.1006/cogp.1993.1012

Alibali, M. W., & Nathan, M. J. (2012). Embodiment in mathematics teaching and learning: Evidence from learners' and teachers' gestures. *Journal of the Learning Sciences*, *21*(2), 247–286. https://doi.org/10.1080/10508406.2011.611446

Alibali, M. W., Nathan, M. J., Fujimori, Y., Stein, N., & Raudenbush, S. (2011). Gestures in the mathematics classroom: What's the point? In N. Stein & S. Raudenbush (Eds.), *Developmental Cognitive Science Goes to School* (pp. 219–234). Routledge, Taylor & Francis.

Atit, K., Gagnier, K., & Shipley, T. F. (2014). Student gestures aid penetrative thinking. *Journal of Geoscience Education*, *63*(1), 66–72. https://doi.org/10.5408/14-008.1

Beattie, G., & Shovelton, H. (1999). Do iconic hand gestures really contribute anything to the semantic information conveyed by speech?: An experimental investigation. *Semiotica*, *123*(1–2), 1–30. https://doi.org/10.1515/semi.1999.123.1-2.1

Beaudoin-Ryan, L., & Goldin-Meadow, S. (2014). Teaching moral reasoning through gesture. *Developmental Science*, *17*(6), 984–990. https://doi.org/10.1111/desc.12180

Becvar, L. A., Hollan, J., & Hutchins, E. (2005). Hands as molecules: Representational gestures used for developing theory in a scientific laboratory. *Semiotica*, *2005*(156), 89–112. https://doi.org/10.1515/semi.2005.2005.156.89
Beilock, S. L., & Goldin-Meadow, S. (2010). Gesture changes thought by grounding it in action. *Psychological Science*, *21*(11), 1605–1610. https://doi.org/10.1177/0956797610385353
Broaders, S. C., Cook, S. W., Mitchell, Z., & Goldin-Meadow, S. (2007). Making children gesture brings out implicit knowledge and leads to learning. *Journal of Experimental Psychology: General*, *136*(4), 539–550. https://doi.org/10.1037/0096-3445.136.4.539
Campbell, J. L., Quincy, C., Osserman, J., & Pedersen, O. K. (2013). Coding in-depth semistructured interviews: Problems of unitization and intercoder reliability and agreement. *Sociological Methods & Research*, *43*(3), 294–320. https://doi.org/10.1177/0049124113500475
Chen, X. (2013). *STEM attrition: College students' paths into and out of STEM fields*. Department of Education, Institute of Education Sciences, National Center for Education Statistics.
Chu, M., & Kita, S. (2011). The nature of the beneficial role of spontaneous gesture in spatial problem solving. *Journal of Experimental Psychology: General*, *140*(1), 102–116. https://doi.org/10.1037/a0021790
Church, R. B. (1999). Using gesture and speech to capture transitions in learning. *Cognitive Development*, *14*(2), 313–342. https://doi.org/10.1016/S0885-2014(99)00007-6
Church, R. B., Ayman-Nolley, S., & Mahootian, S. (2004). The role of gesture in bilingual education: Does gesture enhance learning? *International Journal of Bilingual Education*, *7*(4), 303–319. https://doi.org/10.1080/13670050408667815
Church, R. B., & Goldin-Meadow, S. (1986). The mismatch between gesture and speech as an index of transitional knowledge. *Cognition*, *23*(1), 43–71. https://doi.org/10.1016/0010-0277(86)90053-3
Church, R. B., Schonert-Reichl, K., Goodman, N., Kelly, S. D., & Ayman-Nolley, S. (1995). The role of gesture and speech in communication as reflections of cognitive understanding. *Journal of Contemporary Legal Issues*, *6*, 123–154.
Cook, S. W., & Tanenhaus, M. K. (2009). Embodied understanding: Speakers' gestures affect listeners' actions. *Cognition*, *113*(1), 98–104. https://doi.org/10.1016/j.cognition.2009.06.006
Crowder, E. M., & Newman, D. (1993). Telling what they know: The role of gesture and language in children's science explanations. *Pragmatics and Cognition*, *1*(2), 341–376. https://doi.org/10.1075/pc.1.2.06cro
Ehrlich, S. B., Levine, S. C., & Goldin-Meadow, S. (2006). The importance of gesture in children's spatial reasoning. *Developmental Psychology*, *42*(6), 1259–1268. https://doi.org/10.1037/0012-1649.42.6.1259
Garber, P., & Goldin-Meadow, S. (2002). Gesture offers insight into problem-solving in children and adults. *Cognitive Science*, *26*(6), 817–831. https://doi.org/10.1207/s15516709cog2606_5
Goldin-Meadow, S. (2003). *Hearing gesture: How our hands help us think*. Harvard University Press. https://doi.org/10.1037/e413812005-377
Goldin-Meadow, S., & Alibali, M. W. (2002). Looking at the hands through time: A microgenetic perspective on learning and instruction. In N. Granott & J. Parziale (Eds.), *Microdevelopment: Transition processes in development and learning* (Cambridge Studies in Cognitive and Perceptual Development, pp. 80–106). Cambridge University Press. https://doi.org/10.1017/CBO9780511489709.004
Goldin-Meadow, S., Alibali, M. W., & Church, R. B. (1993). Transitions in concept acquisition: Using the hand to read the mind. *Psychological Review*, *100*(2), 279–297. https://doi.org/10.1037/0033-295X.100.2.279
Goldin-Meadow, S., & Beilock, S. L. (2010). Action's influence on thought: The case of gesture. *Perspectives on Psychological Science*, *5*(6), 664–674. https://doi.org/10.1177/1745691610388764
Goldin-Meadow, S., & Sandhofer, C. M. (1999). Gesture convey substantive information about a child's thoughts to ordinary listeners. *Developmental Science*, *2*(1), 67–74. https://doi.org/10.1111/1467-7687.00056
Goldin-Meadow, S., Wein, D., & Chang, C. (1992). Assessing knowledge through gesture: Using children's hands to read their minds. *Cognition and Instruction*, *9*(3), 201–219. https://doi.org/10.1207/s1532690xci0903_2
Goodwin, C. (2000). Action and embodiment within situated human interaction. *Journal of Pragmatics*, *32*(10), 1489–1522. https://doi.org/10.1016/S0378-2166(99)00096-X
Goodwin, C. (2007). Participation, stance, and affect in the organization of activities. *Discourse & Society*, *18*(1), 53–73. https://doi.org/10.1177/0957926507069457
Goodwin, C. (2010). Things and their embodied environments. In L. Malafouris & C. Renfrew (Eds.), *The cognitive life of things: Recasting the boundaries of the mind,*103-120. McDonald Institute for Anthropological Research.
Hostetter, A. B., & Alibali, M. W. (2008). Visible embodiment: Gesures as simulated action. *Psychonomic Bulletin & Review*, *15*(3), 495–514. https://doi.org/10.3758/PBR.15.3.495
Kelly, S. D., Singer, M., Hicks, J., & Goldin-Meadow, S. (2002). A helping hand in assessing children's knowledge: Instructing adults to attend to gesture. *Cognition and Instruction*, *20*(1), 1–26. https://doi.org/10.1207/S1532690XCI2001_1
McNeill, D. (1992). *Hand and mind: What gestures reveal about thought*. University of Chicago Press.
McNeill, D., Cassell, J., & McCullough, K. E. (1994). Communicative effects of speech-mismatched gestures. *Research on Language & Social Interaction*, *27*(3), 223–237. https://doi.org/10.1207/s15327973rlsi2703_4
Morrell-Samuels, P., & Krauss, R. M. (1992). Word familiarity predicts temporal asynchony of hand gestures and speech. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *18*(2), 615–622. https://doi.org/10.1037/0278-7393.18.3.615

O'Connor, C., & Joffe, H. (2020). Intercoder reliability in qualitative research: Debates and practical guidelines. *International Journal of Qualitative Methods*, *19*, 1–13. https://doi.org/10.1177/1609406919899220
Perry, M., Church, R. B., & Goldin-Meadow, S. (1988). Transitional knowledge in the acquisition of concepts. *Cognitive Development*, *3*(4), 359–400. https://doi.org/10.1016/0885-2014(88)90021-4
Perry, M., & Elder, A. D. (1997). Knowledge in transition: Adults' developing understanding of a principle of physical causality. *Cognitive Development*, *12*(1), 131–157. https://doi.org/10.1016/S0885-2014(97)90033-2
Piaget, J. (1952). *The origins of intelligence in children* (M. Cook, Trans.). W W Norton & Co. https://doi.org/10.1037/11494-000
Pine, K. J., Lufkin, N., & Messer, D. J. (2004). More gestures than answers: Children learning about balance. *Developmental Psychology*, *40*(6), 1059–1067. https://doi.org/10.1037/0012-1649.40.6.1059
Ping, R. M., Goldin-Meadow, S., & Beilock, S. L. (2014). Understanding gesture: Is the listener's motor system involved? *Journal of Experimental Psychology: General*, *143*(1), 195–204. https://doi.org/10.1037/a0032246
Radinsky, J., Ping, R., Hospelhorn, E., & Goldman, S. (2012). Making the absent present: Improvised representational fields in students' negotiation of meaning with GIS. In *Proceedings of the annual meeting of the American Educational Research Association*. AERA.
Rauscher, F. H., Krauss, R. M., & Chen, Y. (1996). Gesture, speech, and lexical access: The role of lexical movements in speech production. *Psychological Science*, *7*(4), 226–231. https://doi.org/10.1111/j.1467-9280.1996.tb00364.x
Rueckert, L., Church, R. B., Avila, A., & Trejo, T. (2017). Gesture enhances learning of a complex statistical concept. *Cognitive Research: Principles and Implications*, *2*(1), 2. https://doi.org/10.1186/s41235-016-0036-1
Siegler, R. S. (2007). Cognitive variability. *Developmental Science*, *10*(1), 104–109. https://doi.org/10.1111/j.1467-7687.2007.00571.x
Singer, M., Radinsky, J., & Goldman, S. R. (2008). The role of gesture in meaning construction. *Discourse Processes*, *45*(4–5), 365–386. https://doi.org/10.1080/01638530802145601
Stieff, M. (2007). Mental rotation and diagrammatic reasoning in science. *Learning and Instruction*, *17*(2), 219–234. https://doi.org/10.1016/j.learninstruc.2007.01.012
Stieff, M. (2011). When is a molecule three dimensional? A task-specific role for imagistic reasoning in advanced chemistry. *Science Education*, *95*(2), 310–336. https://doi.org/10.1002/sce.20427
Stieff, M., & Raje, S. (2010). Expertise algorithmic and imagistic problem solving strategies in advanced chemistry. *Spatial Cognition & Computation*, *10*(1), 53–81. https://doi.org/10.1080/13875860903453332
Trafton, J. G., Trickett, S. B., Stitzlein, C. A., Saner, L., & Schunn, C. D. (2006). The relationship between spatial transformations and iconic gestures. *Spatial Cognition & Computation*, *6*(1), 1–29. https://doi.org/10.1207/s15427633scc0601_1