COGNITIVE SCIENCE
A Multidisciplinary Journal

Cognitive Science 46 (2022) e13180

ISSN: 1551-6709 online
DOI: 10.1111/cogs.13180

The Effect of Cognitive Relevance of Directed Actions on Mathematical Reasoning

Candace Walkington,[a] Mitchell J. Nathan,[b] Min Wang,[a] Kelsey Schenck[b]

[a]*Department of Teaching and Learning, Southern Methodist University*
[b]*Department of Educational Psychology, University of Wisconsin – Madison*

Received 18 May 2021; received in revised form 18 May 2022; accepted 4 July 2022

Abstract

Theories of grounded and embodied cognition offer a range of accounts of how reasoning and body-based processes are related to each other. To advance theories of grounded and embodied cognition, we explore the *cognitive relevance* of particular body states to associated math concepts. We test competing models of action-cognition transduction to investigate the cognitive relevance of directed actions to students' mathematical reasoning in the area of geometry. The hypotheses we test include (1) that cognitively relevant directed actions have a direct effect on performance (direct cognitive relevance hypothesis), (2) that cognitively relevant directed actions lead to more frequent production of gestures during explanations, which leads to improved performance (mediated cognitive relevance hypothesis), and (3) that performance effects of directed actions are influenced by the presence or absence of gesture production during mathematical explanations (moderated cognitive relevance hypothesis). We explore these hypotheses in an experiment where high school students ($N = 85$) evaluated the truth of geometry conjectures after performing cognitively relevant or cognitively irrelevant directed actions while playing a movement-based video game. Contrary to the direct and mediated cognitive relevance hypotheses, we found no overall differences in performance or gesture production between relevant and irrelevant conditions. Consistent with the moderated cognitive relevance hypothesis, cognitive relevance influenced mathematical performance, as measured by the accuracy of students' intuitions, insights, and the validity of their proofs, provided that students produced certain kinds of gestures during mathematical explanations (i.e., with explanatory gestures as the moderator). Implications for theories of grounded and embodied cognition and the design of embodied forms of educational interventions are discussed.

Keywords: Embodied learning; Embodied cognition; Gesture; Geometry; Cognitive relevance

---

Correspondence should be sent to Candace Walkington, Department of Teaching and Learning, Southern Methodist University, Dallas, 75205, TX, USA. E-mail: cwalkington@smu.edu

1. Introduction

Theories of grounded and embodied cognition offer a range of accounts of how intellectual and body-based processes cohere. As grounded and embodied cognition proliferates, and interventions follow suit, an important question arises: *How do specific body states and movements matter for specific forms of thinking and learning?* We conduct our inquiry in the area of mathematical proof practices. The "romance of mathematics" mythologizes mathematics as a discipline based purely on abstractions and formalisms, disconnected from the body and real-world events or objects, yet somehow real (Lakoff & Núñez, 2000, p. xv). Truth, the mythology goes, depends on proofs based on logical, generalizable chains of reasoning that are seldom relegated to prespecified procedures. Yet, increasingly, evidence suggests that people's mathematical thinking and learning are impacted by their body-based experiences and actions. But there is currently no clear, principled guidance for identifying the embodied experiences that will benefit reasoning about specific mathematical ideas.

We test the effects of cognitive relevance as a way to advance the theoretical contributions of grounded and embodied cognition. *Cognitive relevance* addresses the appropriateness of body states to a concept, and, in reciprocal fashion, of a concept to a set of body states. *Body states* in this framework refers to body poses and movements and can include preparatory motor programs that may be activated even if they are not overtly executed, because, for example, they may be inhibited or restricted. Several theoretical accounts are reviewed in our attempts to describe the cognitive relevance of actions and their implications for prevailing theories of cognition, and for mathematical reasoning and education.

We draw on this framing to investigate competing models of action-cognition transduction (ACT; Nathan, 2017). ACT offers specific predictions for (1) how performing conceptually relevant directed actions can activate specific mathematical concepts, and (2) how the gestures that are used to express mathematical reasoning are influenced by or related to these directed actions. The hypotheses that follow from these predictions are empirically investigated in a study that experimentally manipulates the cognitive relevance of directed actions elicited from high school students while playing a movement-based video game. We analyzed the effects of cognitive relevance on students' mathematical reasoning about geometric proofs and the gestures that students produced while verbally expressing their reasoning during in-game prompts. We evaluate the competing models in light of these findings and consider model improvements that further advance understanding of the role of body-based processes in cognition, and for describing embodied approaches to mathematics learning, teaching, and assessment.

2. Theoretical framework

Research on thinking and learning suggests that experiences that sustain learning are enacted, embedded, extended, and embodied (Nathan, 2021); a set of phenomena that are collectively referred to as "4E" (Menary, 2010; Newen, De Bruin & Gallagher, 2018).

Grounded and embodied cognition has arisen as an especially enticing theory of learning, with a growing body of empirical support for a set of claims that posit how concepts, even ones traditionally thought of as being "abstract," attain meaning in inherently body-based ways (Barsalou, 2008; Glenberg, 1997; Shapiro, 2019; Wilson, 2002). A brief review of empirical findings highlights the breadth of these claims.

Children's actions influence how they form shape categories for objects, but only when the children themselves act on the objects in question, not when they observe others perform the same actions (Smith, 2005). A basic bodily structure like hand dominance influences people's judgments about where to place preferred versus less desirable toys, kinder versus meaner stuffed animals, and positive versus negative ideas (Casasanto & Chrysikou, 2011; Casasanto & Henetz, 2012). Language processes, often contrasted with the nonverbal qualities of actions, are influenced by body-based processes as well. Infants learning to recognize speech do more than adjust their hearing, they also activate their lips and tongue to mimic the mouth movements needed to produce the sounds (Bruderer, Danielson, Kandhadai, & Werker, 2015). Teethers that restrict infants' tongue and lip movements can impair their developing auditory speech perception. Neural imaging data of adults reading words with strong motor associations, such as *kick, lick*, and *pick*, show that seeing these words selectively activates—some say it simulates (Barsalou, 2008)—the same motor regions that are activated when people move their feet, tongue, and fingers, respectively (Pulvermüller, 2005).

The role of gestures in language production, language comprehension, and problem solving provides another important pillar of evidence for the central role of the body in complex reasoning (McNeill, 1992). The gestures people produce with speech contribute semantic and pragmatic content. Gestures simulate action, but gestures are seldom in one-to-one relation with spoken words or meaning (Hostetter & Alibali, 2008, 2019). Here, we differentiate gestures from directed actions. Directed actions are physical poses that learners are explicitly instructed by an outside entity to perform, while gestures are generated by the learners themselves (McNeill, 1992). The gestures children produce can reveal the "leading edge" of their development, demonstrating, for example, their awareness of relevant dimensions in conservation tasks before they can verbalize them (Church & Goldin-Meadow, 1986).

People have a gesture threshold (Hostetter & Alibali, 2008), which is the level of motor activation needed for a mental simulation to be expressed in overt action. This threshold can vary depending on factors, such as the current task demands (e.g., strength of motor activation when processing spatial imagery), individual differences (e.g., level of spatial skills), and situational considerations (e.g., social contexts). Hostetter and Alibali (2007) showed the value of including measures of both verbal and spatial skills when modeling gesture production. They found that those with high spatial visualization skills and low verbal skills had the highest rates of gesture production. They found that differences in phonemic fluency, one of their two verbal measures, were associated with differences in gesture production. Gestures are theorized to assist with packaging ideas for speech production (Alibali, Kita, & Young, 2000). Therefore, gesture production may be highest when speakers who have difficulty with phonemic fluency are presented with an organizationally demanding task that draws on gesture support.

Gestures, it is theorized, can also do more than communicate and simulate actions. Gestures can add information from actions into the cognitive encoding of a task. When that information is cognitively relevant to the task, then it can benefit cognitive performance. Explanatory gestures may act as a *mediator* for the effects of directed actions on performance, where the directed actions change the frequency or type of gestures that are generated during explanation, which causes subsequent changes in reasoning. The latter hypothesis builds on research showing that directed actions can influence on the types of gestures learners make during explanations that follow (Cook & Tanehaus, 2009; Donovan & Alibali, 2018). As we will expand on later in the Discussion, learners can make action plans to perform directed actions, which form the basis of motor programs. They also create action plans for their explanatory gestures. When the action plans for explanations and directed actions are in alignment, as can be the case when the directed actions are cognitively relevant to the task they are explaining, they confer an advantage for learners to identify and articulate the relevant generalizable properties of mathematical objects.

This brief review highlights some of the evidence for grounded and embodied cognition. Barsalou (2008, p. 623) observes there is compelling evidence that "increasingly suggest that simulations, situations, and bodily states play central roles in cognition." *Simulations*, in this case, refer to "the reenactment of perceptual, motor, and introspective states acquired during experience with the world, body, and mind" (Barsalou, 2008, p. 618). The body can serve as a mediator for *offline* cognition; that is, for performing intellectual processes even when the target task is not physically present and task-specific actions are not overtly executed.

2.1. General grounded and embodied cognition theories and interventions

Theoretical frameworks have emerged that offer explanatory accounts for how cognitive and body-based motor and perceptual processes interact to realize grounded and embodied cognition. Casasanto and de Bruin (2019) have proposed that directed actions derived from metaphors can enhance word learning through *metaphor congruency*, the consistent mapping of word meaning to actions. Barsalou (1999) proposed that mental processes operate with perceptual symbols, which offers a sensory-motor account of mental representations. Central to this is the idea that thinking about a particular concept involves activating a *perceptual-motor simulation* of the properties associated with that concept, even when no exemplar of the concept is present in the current environment (Barsalou, 1999). Metaphor congruency (Casasanto & de Bruin, 2019) describes how specific actions relate to specific, isolated, concepts, such as *good*, while perceptual symbols (Barsalou, 1999) explain how encountering isolated concepts invoke perceptual-motor simulation processes.

The Indexical Hypothesis (Glenberg & Robertson, 1999) has been proposed to explain why the actions readers perform influence their comprehension and meaning-making processes. Actions help readers *index* the symbols to graspable objects and body-based movements. Glenberg, Gutierrez, Levin, Japuntich, and Kaschak (2004; Adams, Glenberg, & Restrepo, 2019) show significant benefits for sentence and story comprehension for early readers who are directed to touch and manipulate toys in ways that are congruent with the objects and events in the story. These benefits to grounded and embodied reading comprehension extend to performance on mathematical problem solving (Glenberg, Willford, Gibson, Goldberg, & Zhu, 2012). The Indexical Hypothesis extends grounded and embodied cognition in three important ways. First, this research demonstrates ways directed actions can be used as a reliable intervention to facilitate future reasoning, thereby showing causality of the reciprocal relationship between conceptual and perceptual-motor processing. Second, the research extends the influence of grounded and embodied cognition to much richer stimuli, showing its applicability beyond isolated words and concepts to full sentences and stories. Third, it demonstrates that reasoning through carefully selected movements can apply to offline cognition (i.e., imagination), when the task-specific stimuli are not immediately present.

2.2. *Action-cognition transduction*

The bidirectional relationship between body states and cognitive states has been in evidence in a number of experiments. Thomas and Lleras (2007, 2009) conducted studies that provide evidence that body states can influence cognitive states. Participants performed directed arm or eye gaze movements that were either congruent or incongruent to the solutions of otherwise low-performance insight problems. Participants who executed task-relevant directed actions performed better than those performing a variety of task-irrelevant movements, or no movements at all. Notably, participants had these performance benefits without any reported awareness of the relationship between their movements and the problems they solved. Goldin-Meadow, Cook, and Mitchell (2009) found similar results for children directed to make gestures on arithmetic equations that showed equality relationships. This relationship was also demonstrated in results on learning from a scientific text reported by Nathan and Martinez (2015), who manipulated readers' gesture production. They found that inhibiting gesture production selectively impaired inference making during a learning assessment but not surface level recall and general knowledge items, a finding consistent with other research (e.g., Cutica & Bucciarelli, 2008). Nathan and Martinez (2015) proposed that cognitive processes engaged during inference making activated anticipatory simulated actions that triggered overt gesture production (per Hostetter & Alibali, 2008), and that restricting engagement of those body-based resources blocked the anticipatory simulated actions, which impaired respondents' inference-making processes.

ACT has been proposed as a model for describing the bidirectional relationship between body states and cognitive states, and for generating educational interventions that elicit age- and content-appropriate forms of reasoning (Nathan et al., 2014). Nathan et al. (2014) explicitly explored this other half of bidirectionality between body states and cognitive states by investigating how directing people's goal-directed actions influenced their cognitive performance. They found that participants who were directed to perform task-relevant arm movements for a geometry conjecture on triangles and for a parity conjecture involving a chain of gears were more likely to show an understanding of key ideas behind each conjecture, compared to participants who performed motorically similar task-irrelevant movements. As with other studies (e.g., Thomas & Lleras, 2007), those who enjoyed the advantages of task-relevant actions reported not being consciously aware that the motions they performed had anything to do with the conjectures they read.

The image of grounded and embodied cognition that emerges across these frameworks is that there is a reliable, causal, and potentially constitutive relation between one's cognitive state and body state, and that the relation appears to be bidirectional (Shapiro, 2019). There is often a semantic congruency between body movements and the associated reasoning. However, empirical studies of directed actions and gesture production suggest that the relation is not likely to be one-to-one, meaning we cannot expect that there is a unique body (or cognitive) state that reliably induces the corresponding cognitive (or body) state. Indeed, this relation can occur between either concrete or imagined concepts, and either real or simulated actions. Perceptual-motor simulation is implicated as a mechanism for this relation. Cognitive states and perceptual orientations facilitating the enactment of movement appear to be intimately linked—some even propose they are a priori one and the same (e.g., Abrahamson & Sánchez-García, 2016). Conceptually relevant directed actions appear to offer effective interventions for changing people's future reasoning.

2.3. Grounded and embodied mathematical cognition

A number of studies offer empirical evidence of the embodied nature of mathematical knowledge and reasoning. Digital gnosia (i.e., finger discrimination) of children predicts future math performance (Fayol, Barrouillet, & Marinthe, 1998; Reeve & Humberstone, 2011). Adults continue to exhibit this grounded symbol–finger association, activating left-hand muscles for small numbers and right-hand muscles for larger numbers. The SNARC effect offers another example demonstrating the well-established link between spatial–numerical associations and math ability (Berch, Foley, Hill, & Ryan, 1999; Toomarian, Meng, & Hubbard, 2019).

Several studies have shown that activities that leverage the embodied nature of mathematical thinking can improve reasoning and learning. Abrahamson and Bakker (2016) demonstrated ways that eliciting a child's movements that embodied proportionality (such as raising one's hands at rates in ratio of 2 to 3) can serve as the sensorimotor basis for learning multiplicative reasoning. In this line of research, the influences of movement on mathematical reasoning are framed in terms of coordination dynamics. Coordination dynamics is an enactivist account that offers a postcognitivist view of mathematical activity in terms of the regulation of dynamical systems for sensorimotor activity (Abrahamson & Sánchez-García, 2016). When a child performs appropriately to the tasks guided by the Mathematics Imagery Trainer for Proportion digital learning environment, the child has effectively solved an interactive motor-control problem, and in doing so constructed new perceptual structures (denoted as "attentional anchors") that instantiate the targeted mathematical concept (e.g., the ratio of 2 to 3). The subsequent gestures a child generates "demonstrate their productive struggle to coordinate between features of the situation and elements of mathematical forms (Abrahamson, 2004; Nemirovsky, Ferrara, Ferrari, & Adamuz-Povedano, 2020)" (Abrahamson et al., 2020, p. 15).

Studies have also used gesture-based interventions to improve elementary students' understanding of the equivalence relation in arithmetic. In one, children were more likely to perform and verbally describe a correct, novel strategy for solving equivalent equations when directed to gesture in ways that highlighted the two sides of the equation, coming to overtly express a strategy that had previously been only implicit (Goldin-Meadow et al., 2009). Research has also shown it is not simply all actions that contribute to generalizable learning in mathematics. Children performing concrete actions showed poorly developed understanding of the verbal principle of maintaining equal values on both sides of the equation, while those taught to perform schematized gestures that stripped away the particulars of any one instance showed the greatest transfer (Novack, Congdon, Hemani-Lopez, & Goldin-Meadow, 2014).

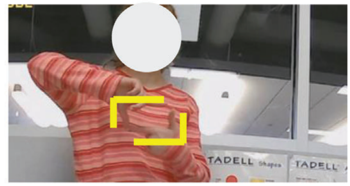
Directed actions also have been used to influence people's statistical reasoning. Zhang, Givvin, Sipple, Son, and Stigler (2021) asked participants to perform a secondary task by moving their hands to track placement and orientation of rectangles that were overlaid atop an instructional video teaching the concept of using equations as a statistically predictive model. Those performing hand movements designed to be conceptually relevant to the mathematics outperformed those whose movements were conceptually irrelevant, as well as those who did no directed actions. The work is especially relevant to the current exploration because it manipulated the conceptual relevance of the actions performed and did so without overtly communicating the relevance of the actions to the participants.

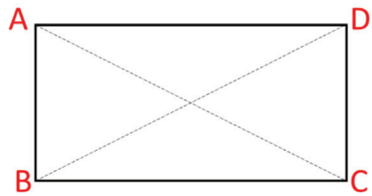
2.4. Proving geometry conjectures: Intuition, insight, and proof

Geometry in secondary education focuses on developing learners' skills in evaluating the universal truth values of conjectures about space and shape and justifying their conclusions through the production of mathematical proofs. Proof is the primary methodology for generating new mathematical knowledge (Rav, 1999). We align ourselves with Harel and Sowder (1998, p. 241), who define proving as "the process employed by an individual to remove or create doubts about the truth of an observation." *Transformational* proofs enlist mental or physical operations to demonstrate the validity and generality of conjectures (Clements & Battista, 1992). As a form of deductive analytic proof (Harel & Sowder, 2007), valid transformational proofs must conform to three defining characteristics: They must be *general*, showing the argument is true for all members of an object class; use *operational thought*, where the prover progresses systematically through a goal structure, anticipating the outcomes of proposed transformations; and follow a chain of *logical inference*, with conclusions following from valid premises.

Empirical studies of professional mathematicians engaged in proof practices reveal "that gesture and other bodily movement is essential … in the intellectual construction of mathematics;" and along with words, symbols, diagrams, and objects, "the mathematician's body may be a constitutive part of his or her situated proving" (Marghetis & Núñez, 2013, p. 229). Similarly, students engaged in proofs regularly produce gestures (Kim, Roth, & Thom, 2011; Marghetis, Edwards, & Núñez, 2014; Nathan et al., 2021; Ng & Sinclair, 2015a; Pier et al., 2019; Williams-Pierce et al., 2017). For our purposes, a *gesture* is defined as a spontaneous or purposeful body movement that conveys meaning, and that often accompanies speech or thought (Goldin-Meadow, 2003; Kita, Alibali, & Chu, 2017; McNeill, 1992; Walkington, Chelule, Woods, & Nathan, 2019). Fig. 1 displays a student using gestures to reason through

"I guess it just means whenever you get like a rectangle and like split it, are the same."
(use thumbs and index fingers to form a rectangle)
(point at the mirrored right angle with index finger).
"Cause um, cause the short side and the long side are like they're equal and the same and also Pythagorean Theorem, you know 'a' squared plus 'b' squared equals 'c' squared."
(uses thumbs and index fingers to form a rectangle)





Statements	Reasons
1. ABCD is a rectangle.	1. Given
2. $\overline{AB} \cong \overline{CD}$	2. Definition of rectangle
3. $\overline{BC} \cong \overline{BC}$	3. Reflexive property
4. m $\angle ABC$ = m $\angle BCD$=90°	4. Definition of rectangle
5. $AB^2+BC^2=AC^2$ and $CD^2+BC^2=BD^2$	5. Pythagorean theorem
6. $\overline{AC} \cong \overline{BD}$	6. Properties of equality

Fig. 1. Transformational proof (top) and traditional two column proof (bottom) for the conjecture "The diagonals of a rectangle are always congruent."

a conjecture about the diagonals of any rectangle being congruent, engaging in a transformational proof. Learners regularly use gestures in the mathematics classroom (Nathan, Alibali, & Church, 2017), and gesture use among learners is correlated with stronger mathematical reasoning (Cook & Goldin-Meadow, 2006; Goldin-Meadow, 2003; Nathan et al., 2021; Pier et al., 2019).

Gestures have affordances and constraints for mathematical representation and communication when compared to other media like screens with dynamic geometry software (DGS) or paper. Gestures, along with being dynamic, are extremely portable and flexible, have low barriers to entry, and are usually meaningful to the gesturer. Writing also has these affordances (but is relatively nondynamic), along with the advantage of a visible record of thinking being maintained. DGSs can constrain the level of interactivity of mathematical objects, preventing task-irrelevant manipulations (Barrett, Stull, Hsu, & Hegarty, 2015) and allowing students to dynamically view and interact with invariant properties of objects and take precise measurements, which instantaneously update as the object is changed (Hollebrands, 2007). Some

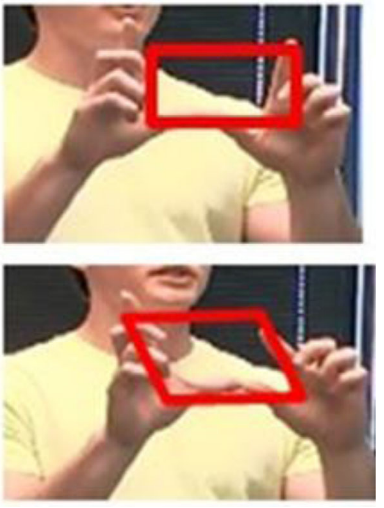


Fig. 2. Participant makes dynamic gesture of sides of a rectangle being folded to make a parallelogram while proving conjecture 5 in Fig. 4. Red outlines added for emphasis.

DGSs can induce a high cognitive load when learning the interface (Reis et al., 2012), while some more transparent user interfaces may prevent formal mathematical operations—like rotation—from being consciously recognized in the way they would if students were using menus (Sarama & Clements, 2009).

Pier et al. (2019) showed that gesture production during geometry proof was a reliable predictor of participants' mathematical insights and likelihood of generating mathematically valid proofs, even when controlling for spatial ability, verbal fluency, prior geometry knowledge, gender, and English language proficiency. The investigators specifically identified dynamic gestures as instrumental for valid, generalizable reasoning. *Dynamic gestures* depict motion-based transformations of perceived mathematical objects as they are transformed through multiple states, such as rotation, reflection, dilation, and skewing. An example of a dynamic gesture (Fig. 2) is formulating a parallelogram with one's hands, and then skewing the sides to explore whether its area is preserved. This could be contrasted with a nondynamic gesture, where the learner might simply form the parallelogram with their hands or even trace its static form, but not modify or transform it in any way. Dynamic gestures have been previously implicated for superior reasoning in mathematics (Garcia & Infante, 2012; Pier et al., 2019) and physical science (Göksun, Goldin-Meadow, Newcombe, & Shipley, 2013). The nature of dynamic gestures is closely bound to the mathematical task that they enact, and always involve the representation and transformation of the specific mathematical objects and features contained in the task. Spatial reasoning is also an important component of geometric reasoning (Jones & Tzekaki, 2016; Sinclair et al., 2016) and strong associations have been found between spatial reasoning, dynamic gestures, and valid geometric reasoning and proof (Göksun et al., 2013; Nathan et al., 2021).

2.5. The proposed models, hypotheses, and research questions

Previous grounded and embodied cognition research provides evidence that concepts can be grounded in body-based action, and that performing actions can have a reciprocal effect on reasoning processes. Scholars have demonstrated that interventions can be designed to invoke cognitively relevant embodied experiences to enhance performance on specific tasks. The gestures that people produce during task performance also influence one's reasoning processes.

Gestures enact simulated actions invoked by the processes involved when performing mental manipulations; and, as actions themselves, may influence reasoning processes. Directed actions may, therefore, spur learners to make particular kinds of gestures, which in turn may influence learners' reasoning processes. Dynamic gestures have been identified as important for simulating the transformations of physical and mathematical objects. In this view, we conceptualize gesture production as a *mediator* that helps explains the effect of directed actions on cognitive performance. Within this theoretical frame, two important research questions arise.

RQ1. Does the cognitive relevance of directed actions impact students' mathematical performance?

Actions aid cognition. Less clear is how closely these actions need to cohere to the concepts under investigation. We investigate this question in order to understand whether embodied interventions can be effectively designed to improve mathematical reasoning. The *direct cognitive relevance hypothesis* is that participants who engage in directed actions that are cognitively relevant to the mathematical concepts under investigation will show superior performance as assessed by their mathematical intuition, insight, and proof validity, compared to those who perform cognitively irrelevant actions.

RQ2. Does the cognitive relevance of the directed actions impact students' tendency to gesture?

Students who produce gestures, especially dynamic gestures, exhibit superior mathematics performance. Because dynamic gestures simulate transformations of mathematical objects given in the task, they are constrained to gestures of mathematical objects and manipulations that are relevant to the task. Because of this, we would not expect dynamic gestures to be facilitated by cognitively irrelevant directed actions. This suggests that dynamic gestures may play a mediational role in the construction of one's mathematical reasoning. We explicitly investigate the *mediated cognitive relevance hypothesis*, the proposition that cognitively relevant directed actions lead to more frequent production of gestures, which in turn leads to improved cognitive performance. This mediational relationship is illustrated by the model shown in Fig. 3 (top).

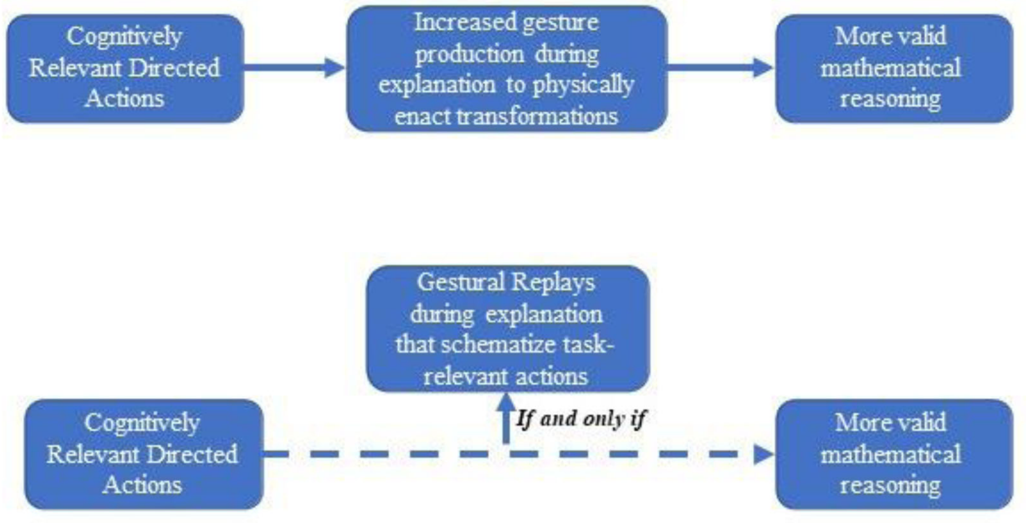


Fig. 3. Visual depiction of mediated cognitive relevance hypothesis (top) and moderated cognitive relevance hypothesis (bottom).

3. Methods

3.1. Participants

Participants included $N = 85$ high school students from eight high schools in a metropolitan area. The students were enrolled in a support program for high school students poised to become first-generation college students. Typically, all students in this program qualified for free/reduced lunch. Sixty-five were females and 20 were males; 25 were in 9th grade, 29 were in 10th grade, 22 were in 11th grade, and 9 were in 12th grade. In addition, 48 identified as Hispanic, 26 as African-American, 5 as Asian, 3 as Caucasian, 2 as Other race/ethnicity, and 1 had race/ethnicity data missing. Eleven students had taken Algebra 1 as their current/most recent math class, and thus had not yet taken Geometry. Thirty students had taken Geometry as their most recent math class, 25 had taken Algebra 2 as their most recent math class, 14 had taken Pre-Calculus as their most recent math class, 3 had taken Calculus as their most recent math class, and 2 (in 11th and 12th grades) selected "Other" as their most recent math class. Participants were recruited from a pool of 349 high school students in the support program. Study participants were selected from those who had signed parental consent forms and who were present during data collection days. They received a $20 gift card for their participation.

3.2. Experiment protocol

Participants were video-recorded while individually playing a motion capture game with a laptop attached to a Kinect^{TM} camera called *The Hidden Village*. In the game, on-screen avatars (Fig. 4) would direct players to perform sequences of arm motions, and the Kinect camera would detect whether players successfully made the motions and advance them

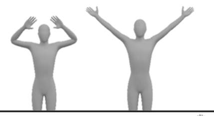
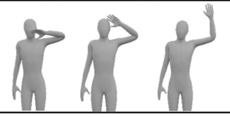
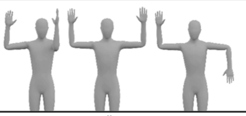
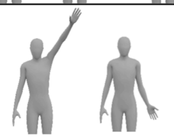
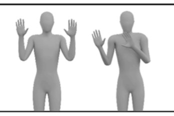
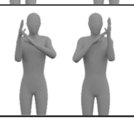
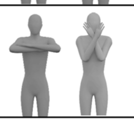
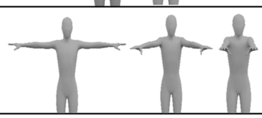
Conjecture	Directed Actions	Intended Relevance
1. Given that you know the measure of all three angles of a triangle, there is only one unique triangle that can be formed with these three angle measurements. (False)		Poses show a triangle getting larger.
2. If one angle of a triangle is larger than the second angle, then the side opposite first angle is longer than the side opposite the second angle.		Poses show one angle of a triangle opening up and getting larger, as opposite side also gets larger.
3. If you double the length and the width of a rectangle, then the area is exactly doubled. (False)		Poses show doubling the length of the rectangle, and then further doubling its height.
4. Reflecting any point over the x-axis is the same as rotating the point 90 degrees clockwise about the origin. (False)		Poses show a point being reflected over the x axis (horizontial at shoulder) and the rotation being greater than 90 degrees.
5. The area of a parallelogram is the same as the area of a rectangle with the same base and height.		Poses show a rectangle, and then a parallelogram where area is re-organized.
6. The diagonals of a rectangle always have the same length.		Poses show two right triangles that make up the diagonals of a recntagle.
7. The opposite angles of two lines that cross are always the same.		Poses show sets of vertical angles that are forced to be equal no matter how two lines are positioned.
8. The sum of the length of two sides of a triangle is always greater than the length of the third side.		Poses show possible and impossible triangles.

Fig. 4. Relevant action sequences for eight conjectures. Irrelevant action sequences are not shown, but involved the same total number of poses being performed. These poses were selected from the total set of relevant action poses.

in the game if they did. These motions were designed to either be relevant or irrelevant to subsequent geometry conjectures that students would be asked to prove (Fig. 4). For example, a player might be asked to prove the conjecture "The diagonals of a rectangle always have the same length," and before that conjecture might perform relevant motions of two congruent mirror-images of right triangles (see conjecture 6 in Fig. 4). The right triangle motions are intended to convey a key insight related to proving the conjecture—that the diagonals must be congruent because they each are the hypotenuses of right triangles that have equal leg lengths.

The directed actions given by the game were determined by examining the spontaneous hand and arm gestures that successful problem-solvers tended to make when proving these conjectures (see Nathan et al., 2020; Walkington, Woods, Nathan, Chelule, & Wang, 2019).

These hand gestures were then adjusted to meet the affordances and limitations of the motion detection technology, which could only detect arm movements. For example, Fig. 2 shows an image of the gestures a participant used to prove the conjecture "An area of a parallelogram is the same as the area of a rectangle with the same base and height." The participant gestured a rectangle being transformed into a skewed parallelogram, with its area rearranged but still equal, using thumbs and index fingers to form the shapes (red outlines were added to the images). These gestures were then reimagined for use in the game as the arm poses shown in conjecture 5 in Fig. 4.

Each participant performed such relevant motions for four of the eight conjectures they proved and performed irrelevant motions for the remaining four. The irrelevant motions were the exact same set of arm poses used in the relevant motions shown in Fig. 4, but the arm poses were in different orders (e.g., the player might perform pose #2 from conjecture 1 and then pose #3 from conjecture 3 as an irrelevant motion sequence when they were proving conjecture 4). A Latin square design was used to account for any ordering effects of conjecture. Performing relevant and irrelevant motions was alternated for each associated conjecture, with the type of motions given for the first conjecture counterbalanced. Participants were not informed that any of their arm motions were relevant to the conjectures they were being asked to prove. For participants, the motions were imbedded in the game mechanics and were necessary to move forward in the gameplay sequence.

After performing one set of relevant or irrelevant motions, participants were next asked to read the associated conjecture out loud and explain why the conjecture was true or false out loud. They then advanced to a screen where they needed to choose the best explanation from four multiple choice options of explanations for why the conjecture was true or false. Students' multiple-choice selections were not predicted by experimental condition nor by interaction terms involving experimental condition, so we only consider students' verbal explanations in the analyses reported below. The participant cycled through this process eight times for the eight conjectures.

An interviewer was present at all sessions and followed a script for all interactions with students. Students were instructed that the interviewer could not give them any assistance, other than defining four vocabulary words that were bolded (parallelogram, diagonals, reflection, and rotation). The interviewer gave generic prompts at predetermined points asking the students to read the conjectures out loud to explain out loud why they thought each conjecture was true or false. The interviewer also asked students at the end of the session whether they saw a connection between their directed actions and the mathematical conjectures. The experimental protocol is depicted in Fig. 5.

3.3. Data sources and measures

Participants were first given a Qualtrics survey, which requested demographic information, including their age, race/ethnicity, grade level, most recent math course name and the grade they received, and languages spoken. The survey contained eight items assessing students' individual interest in geometry, using the scales in Linnenbrink-Garcia et al. (2010) (reliability = 0.8; see online Appendix A). Interest is defined as the process of engaging with and

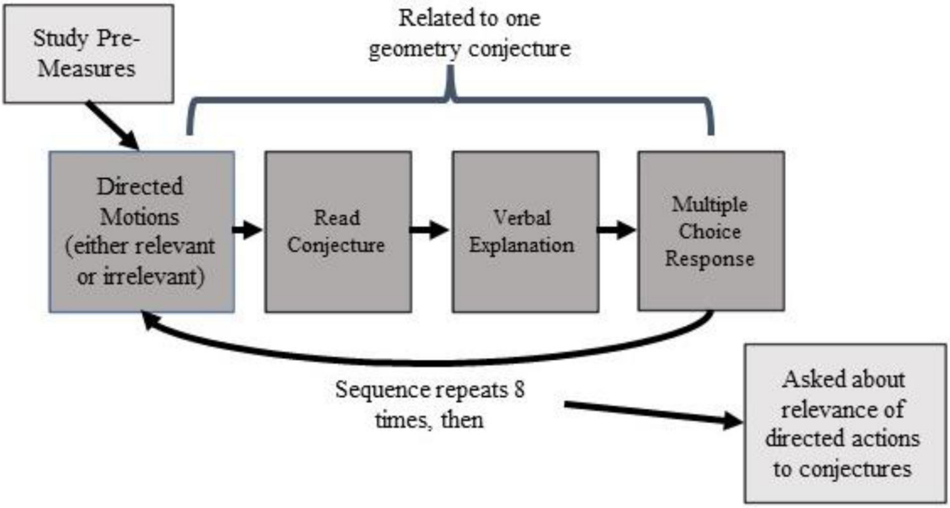


Fig. 5. Flow of experimental protocol.

the predisposition to re-engage with certain topics or ideas (Hidi & Renninger, 2006); the instrument we selected focused on the predisposition to engage with geometry. Interest in mathematics has been shown to be a strong predictor of motivation and achievement (e.g., Kim, Jiang, & Song, 2015; Renninger & Hidi, 2015).

Students were also given a 12-item pretest of their geometry knowledge where they were asked to recall basic properties of triangles, circles, and parallelograms. The pretest did not turn out to be a reliable measure based on its internal consistency, so it was not used in the analyses. Students were then given a timed spatial visualization test where they saw images of holes punched in folded paper and had to guess where the holes would be when the paper was unfolded (Ekstrom, French, Harman, & Derman, 1976; reliability = 0.76). Finally, students were given a phonemic fluency measure where they had to name as many words as they could in 60 s (out loud) that began with “s” and then “t.” (desRosiers & Kavanagh, 1987; reliability = 0.88). Descriptive statistics for all measures are shown in Table 1.

Online Appendix B gives the coding criteria for intuition, insight, and proof for each of the eight conjectures. We follow Zhang, Lei, and Li (2016) in defining intuition as participants’ snap judgments of the verity of conjectures, and insight as their understanding of key mathematical ideas, or “gist,” behind the conjecture. We follow Harel and Sowder’s (2007) previously stated definition of transformational proof, where proofs must be general and involve logical inference and operational thought. Inter-rater reliability was evaluated on a subset of 120 video clips. These clips were coded in six sets of approximately 20 clips each. Approximately 25% of the clips in each set were selected randomly from a subset of clips that the first coder believed showed correct insight or proof. The remainder were randomly selected from the full set. All clips were mixed together when given to the second coder. Selection was carried out in this manner so there would be enough clips with correct insight

Table 1
Descriptive statistics for all measures

Measure	Mean ($n = 85$)	Standard deviation
Geometry Interest (1–5 scale)	3.40	0.70
Spatial Visualization (out of 20)	7.18	4.41
Phonemic Fluency Measure (count of words in 60 s)	10.74	3.50
	Mean	*n*
Intuition Correct	48.85%	656
Insight Correct	19.13%	663
Proof Correct	5.87%	663
Any Gesture	21.76%	657
Depictive Gesture	17.35%	657
Dynamic Gesture	9.44%	657

Note: Standard deviations are not given for the outcome and gesture variables, as they are 0/1 variables. Instead, the sample size is given—although there were a total of 680 conjectures given to students, coding could be missing for a variety of reasons, including the student response being too soft to hear and the camera not being aimed adequately at students' hands to see gestures. The phonemic fluency measure was missing for one student because its administration was not recorded.

and proof to meaningfully obtain inter-rater reliability on these categories. Cohen's kappa was 0.82 for proof, 0.81 for insight, and 1.0 for intuition.

We coded whether participants made any gestures at all while proving each conjecture (not counting nonmathematical gestures like nodding or adjusting one's hair, but including pointing gestures), whether they made any depictive gestures (see McNeill, 1992) where they used their hands to physically form a point, shape, or line, and whether they made any dynamic gestures where they showed the transformation of a mathematical object through multiple states. All gesture categories were coded as 0/1 (present/not present) for the entire clip of one participant proving one conjecture. Present or absent coding was used for the participants' gestures and explanations given the relatively short and simple nature of the explanations—the average number of words in an explanation was 19.3 ($SD = 15.1$). The distribution for how often each gesture category appeared in the data can be found in Table 1. The "any gesture" category included all trials with dynamic gestures, depictive gestures, and pointing gestures; of the 148 instances of any gesture, 114 were depictive and 62 were dynamic. Of the 114 instances of depictive gesture, 36 of those trials were also coded as dynamic.

Inter-rater reliability was evaluated on a subset of 120 video clips. These video clips were the same ones selected for the above inter-rater reliability analysis, and thus were selected in a way that oversampled clips where students were likely to be engaging in correct insights or proofs. Cohen's kappa values were 0.87 for any gesture, 0.74 for dynamic gesture, and 0.73 for depictive gesture.

3.4. Power analysis

Power analyses were conducted apriori to determine the appropriate sample size to detect the effect of the treatment variable (relevant/irrelevant actions) on the most important outcome (mathematical insight). We originally estimated an effect size of $f = 0.31$ for the effect of relevant action on insight performance, which is drawn from Nathan et al. (2014). We used G*Power 3.1.9.2 (Faul, Erdfelder, Buchner, & Lang, 2009) using a power level of 0.80 and $\alpha = 0.05$. Using the ANOVA Repeated Measures-Within with two conditions and a correlation among repeated measures of 0.5, a minimum sample of 24 was needed.

Since our mediation analyses needed to be powered as well, we also examined the guidelines for Baron and Kenny's (1986) mediation test given in Fritz and MacKinnon (2007). Pier et al. (2019) study suggests that the effect size of our mediator, dynamic gesture, on proof performance is at least 0.5, thus our beta path has a medium effect size. The effect size of the intervention on our mediator (i.e., the alpha path) was less established by previous data at the time we were conducting this power analysis in 2015. We used Petrick's (2012) study comparing observing versus enacting geometric relationships and Walkington, Nathan, Wolfgram, Alibali, and Srisurichan (2014) study comparing gesture inhibition to no inhibition to suggest a medium ($d = 0.4$–0.6) effect size. Thus, based on Fritz and MacKinnon (2007), we conservatively aimed for a minimum effective sample size (after taking into account clustering via the design effect) of 125, which can detect small/medium mediation paths of 0.26 (alpha) and 0.39 (beta). Given the design effect of 4.5, we determined that 76 total participants would give us an effective sample size of 135, which accounts for some attrition. We ended up with 85 total participants because we did not stop running participants until the end of the final day that we had scheduled for data collection.

We note that the mediator hypothesis and the direct cognitive relevance hypothesis were the two hypotheses that we had generated based on the literature before the experiment began. When it became clear that our data were not entirely consistent with these hypotheses, we posed a third hypothesis, the moderator hypothesis, as an alternative way to explore the findings.

3.5. Data analysis

Data were analyzed using mixed effects logistic regression models (Snijders & Bosker, 1999) where repeated observations of students solving conjectures were nested within student. Participant ID was modeled as a random effect and an additional random effect was added to the models for which conjecture the participant was proving. Our main outcome variables included (1) whether the participant made the correct judgment about whether the conjecture was true or false, "intuition," (2) whether the participant showed an understanding of key ideas behind the proof, "insight," and (3) whether the participant gave a valid, generalized mathematical proof for the conjecture. For Research Question 2, gesture was the outcome variable. Mixed-effects logistic regression models were fit with participant ID and conjecture as random effects. Student characteristics were tested for significance in the models as covariates (gender, language, math course grade, whether their most recent math class was above or below geometry or was geometry, math interest, phonemic fluency, spatial test score). In our models, each data point was one instance of one participant proving one conjecture, for a total of 85*8 = 680 data points. The directed action Condition (relevant or irrelevant) was the main predictor.

Models were fit using the *glmer* command of the lme4 package in R (Bates, Maechler, Bolker, & Walker, 2014; R Core Team, 2018). This command can handle mixed effects data that are partially crossed, partially nested, and unbalanced. Our mixed effects logistic regression model (Snijders & Bosker, 1999) has the form:

$$logit(P_{ij}) = \gamma_{00} + \gamma_1 \times (experimental\ condition) + \sum_{h=2}^{q} \gamma_h \times (student\ characteristics)_h + U_{0j\ student} + T_{0j\ problem} \quad (1)$$

Our approach was to first fit this model for RQ1. We then proceeded to model our mediator (gesture) as the outcome (RQ2), using the equation:

$$logit(M_{ij}) = \gamma_{00} + \gamma_1 \times (experimental\ condition) + \sum_{h=2}^{q} \gamma_h \times (student\ characteristics)_h + U_{0j\ student} + T_{0j\ problem} \quad (2)$$

Our next step would have been to use Eq. 3 with the mediator variable included (i.e., Kenny, Kashy, & Bolger, 1998). However, given a lack of significant effects in (2), we instead fit (4) where gesture was a moderator as part of an exploratory analysis, where we examined the interaction of gesture with relevant/irrelevant condition.

$$logit(P_{ij}) = \gamma_{00} + \gamma_1 \times (experimental\ condition) + \gamma_2 \times (moderator\ variable) + \sum_{h=2}^{t} \gamma_h \times (student\ characteristics)_h + U_{0j\ student} + T_{0j\ problem} \quad (3)$$

$$logit(P_{ij}) = \gamma_{00} + \gamma_1 \times (experimental\ condition) + \gamma_2 \times (moderator\ variable) + \gamma_3 \times (moderator\ variable) \times (experimental\ condition) + \sum_{h=4}^{t} \gamma_h \times (student\ characteristics)_h + U_{0j\ student} + T_{0j\ problem} \quad (4)$$

Student characteristics were only retained in the model if they were statistically significant. The random effects were retained in the model regardless of significance. Two-way interactions for (4) were tested using likelihood ratio tests on nested models to test for significant reductions in deviance. Effect size was measured using *d*-type measures (see Chinn, 2000), with 95% confidence intervals computed.

Table 2
Mixed effects logistic regression models predicting proof, insight, and intuition based on experimental condition

	Model 1: Proof	Model 2: Insight	Model 3: Intuition
Random Effects: (St. Dev.)			
Participant ID	0	0.52	0.45
Conjecture	0.80	0.44	0.42
Fixed Effects:			
(Intercept)	–3.53 (0.43)	–1.60 (0.23)	–0.01 (0.20)
Irrelevant Actions	(ref.)	(ref.)	(ref.)
Relevant Actions	0.35 (0.35)	–0.31 (0.21)	–0.10 (0.17)
Spatial Test Score (Centered)	0.14 (0.04)***	0.13 (0.03)***	0.10 (0.02)***
Geometry Interest (Centered)	0.69 (0.28)*	0.47 (0.19)*	
Phonemic Fluency (Centered)	0.10 (0.05)*	0.10 (0.04)**	

*= $p < .05$.
**= $p < .01$.
***= $p < .001$.

4. Results

4.1. Research Question 1: Main effects of directed actions on performance

Regression results for RQ1 are given in Table 2. Results showed that spatial test score was a positive predictor of intuition ($d = 0.054$ per point on 20-point test, 95% CI [0.029, 0.078], $p < .001$), insight ($d = 0.069$ per point on test, 95% CI [0.039, 0.100], $p < .001$), and proof ($d = 0.077$ per point on test, 95% CI [0.034, 0.120], $p < .001$). Geometry interest significantly positively predicted insight ($d = 0.261$ per 1 point on a 5-point scale, 95% CI [0.060, 0.463], $p = .011$) and proof ($d = 0.384$ per 1 point on 5-point scale, 95% CI [0.078, 0.689], $p = .014$). In addition, phonemic fluency significantly positively predicted insight ($d = 0.058$ per 1 point, 95% CI [0.019, 0.096], $p = .003$) and proof ($d = 0.054$ per 1 point, 95% CI [0.002, 0.107], $p = .042$). However, as can be seen from Table 2, and as is illustrated in Fig. 6, action condition (relevant vs. irrelevant) was not a significant predictor of intuition, insight, or proof. These results are not consistent with the direct cognitive relevance hypothesis.

4.2. Research Question 2: Main effects of directed actions on gesture

Regression results for RQ2 are given in Table 3. Results showed that spatial test score significantly predicted all three gesture categories—dynamic gesture ($d = 0.09$ per point on 20-point test, 95% CI [0.040, 0.136], $p < .001$), depictive gesture ($d = 0.06$ per 1 point, 95% CI [0.009, 0.119], $p = .023$), and any gesture ($d = 0.07$ per 1 point, 95% CI [0.006, 0.127], $p = .031$). Phonemic fluency also significantly predicted dynamic gesture ($d = 0.06$ per 1 point, 95% CI [0.0007, 0.13], $p = .048$). As can be seen from Table 3, and as is illustrated in Fig. 7, action condition (relevant vs. irrelevant) was not a significant predictor of gesture usage in any of the three categories. These results are not consistent with the mediated cognitive relevance hypothesis. Given the lack of evidence for the mediated cogni-

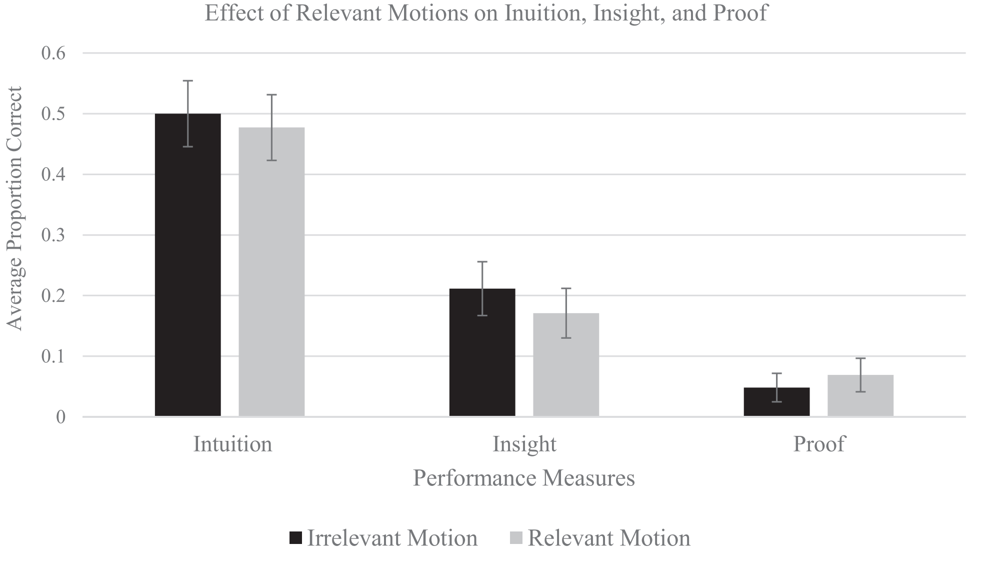


Fig. 6. Main effects of relevant actions on intuition, insight, and proof. Error bars represent standard error of the mean. The sample size for each bar is 85 participants.

Table 3
Mixed effects logistic regression predicting gesture usage based on experimental condition

	Model 4: Dynamic Gesture	Model 5: Depictive Gesture	Model 6: Any Gesture
Random Effects: (St. Dev.)			
Participant ID	0.88	1.55	1.82
Conjecture	1.30	0.62	0.82
Fixed Effects:			
(Intercept)	–3.27 (0.58)	–2.35 (0.37)	–2.08 (0.42)
Irrelevant Actions	(ref.)	(ref.)	(ref.)
Relevant Actions	–0.17 (0.33)	0.03 (0.24)	–0.14 (0.24)
Spatial Test Score (Centered)	0.16 (0.04)**	0.12 (0.05)*	0.12 (0.06)*
Phonemic Fluency	0.11 (0.06)*		

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

tive relevance hypothesis, we move to an exploratory analysis where we instead treat gesture as a moderator.

4.3. *Exploratory analysis: Gesture usage as a moderator of the effect of directed actions*

Above, we reported our preplanned analyses that related to our original hypotheses; next, we will describe exploratory analyses we conducted. Figs. 8–10 show how intuition (Fig. 8), insight (Fig. 9), and proof (Fig. 10) outcomes varied according to action condition

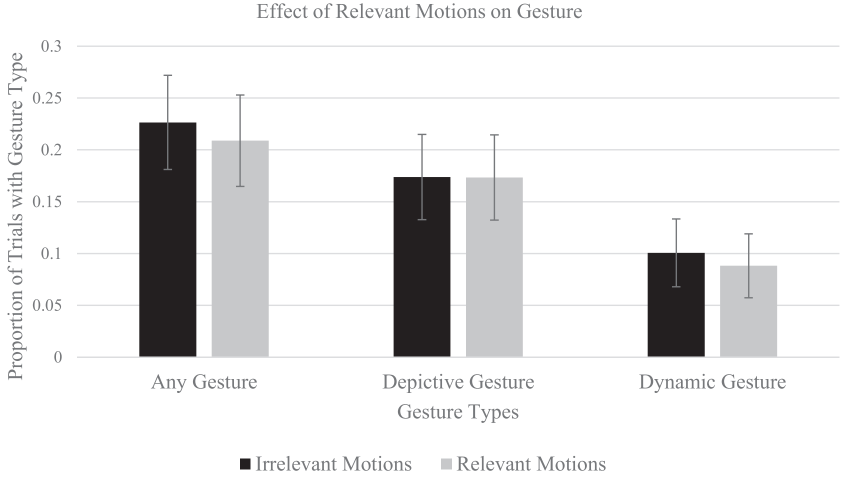


Fig. 7. Main effects of relevant actions on gesture, across three different gesture categories. Error bars represent standard error of the mean. The sample size for each bar is 85 participants.

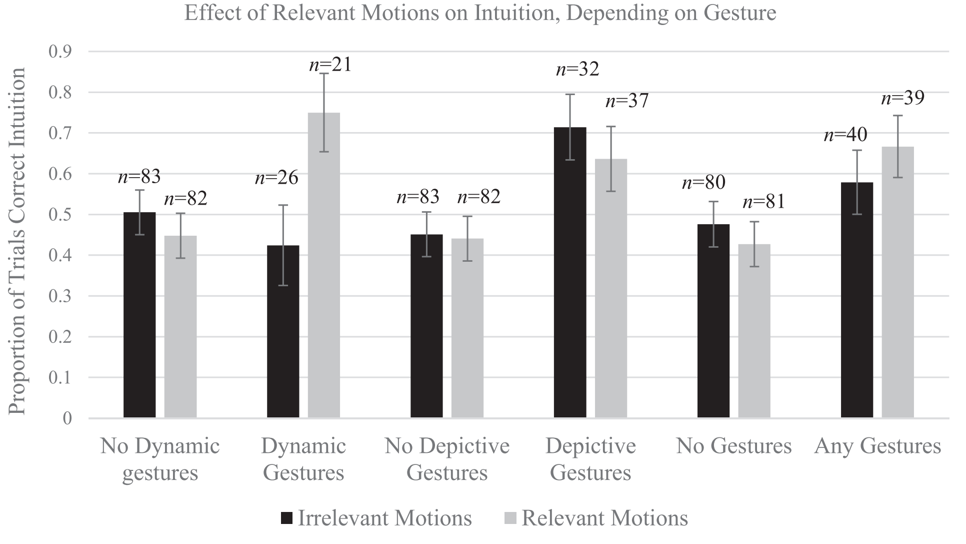


Fig. 8. Rate of Correct Intuition, by whether participants made three different categories of gesture (any gesture, depictive gesture, and dynamic gesture). Error bars represent standard error of the mean. The "*n*" number above each bar gives the sample size that the bar was calculated from, out of the 85 participants. Bars that involve "no gestures" tend to have a higher sample size, because to be excluded from this bar, the participant would have had to gesture in the specified way (or be coded as NA) for all four conjectures. Bars that involve the presence of gestures have smaller sample sizes, as they only include participants who performed at least one gesture of this type (e.g., 26 participants in the irrelevant motions condition produced at least one dynamic gesture). These are the sample sizes that were used to compute the error bars.

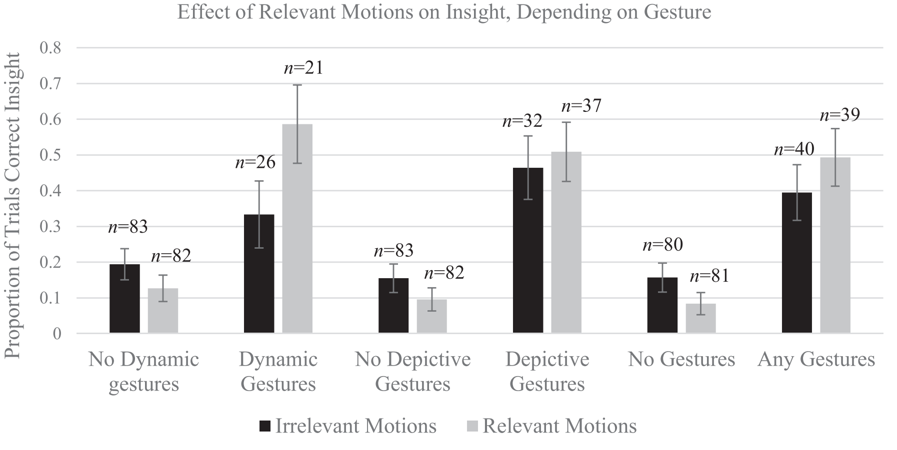


Fig. 9. Rate of Correct Insight, by whether participants made three different categories of gesture (any gesture, depictive gesture, and dynamic gesture). Error bars represent standard error of the mean. The "*n*" number above each bar gives the sample size that the bar was calculated from, out of the 85 participants. Bars that involve "no gestures" tend to have a higher sample size, because to be excluded from this bar, the participant would have have had to gesture in the specified way (or be coded as NA) for all four conjectures. Bars that involve the presence of gestures have smaller sample sizes, as they only include participants who performed at least one gesture of this type (e.g., 26 participants in the irrelevant motions condition produced at least one dynamic gesture). These are the sample sizes that were used to compute the error bars.

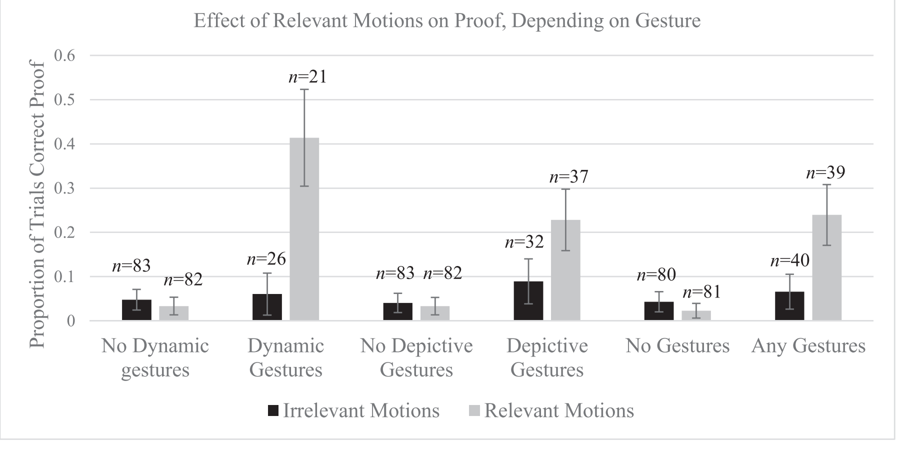


Fig. 10. Rate of Correct Proof, by whether participants made three different categories of gesture (any gesture, depictive gesture, and dynamic gesture). Error bars represent standard error of the mean. The "*n*" number above each bar gives the sample size that the bar was calculated from, out of the 85 participants. Bars that involve "no gestures" tend to have a higher sample size, because to be excluded from this bar, the participant would have have had to gesture in the specified way (or be coded as NA) for all four conjectures. Bars that involve the presence of gestures have smaller sample sizes, as they only include participants who performed at least one gesture of this type (e.g., 26 participants in the irrelevant motions condition produced at least one dynamic gesture). These are the sample sizes that were used to compute the error bars.

and whether or not the participant made gestures that fell into each of the three gesture categories during the trial. For intuition and insight, the graphs suggest that there may not be reliable differences between relevant and irrelevant actions, unless the participant made dynamic gestures during the trial, in which case relevant actions seemed to be associated with higher performance. For proof, we see this pattern for all three gesture categories. When each gesture category is not present, there is little difference between outcomes for relevant and irrelevant action trials; however, when gestures are present, relevant actions seem to be associated with higher proof performance. These data suggested to us that there were some important situations in which directed actions were beneficial for learners, which were not being captured in our original models or hypotheses. We thus fit new models where gesture was a *moderator* (fit as an interaction term) of the effect of directed actions on performance.

Table 4 shows how the results shown in the bar graphs bear out in the regression models. Model selection warranted the inclusion of a Condition × Any Gesture interaction for the proof model ($\chi^2(1) = 9.40$, $p = .002$) and the insight model ($\chi^2(1) = 9.81$, $p = .002$), a Condition × Depictive Gesture interaction for the insight model ($\chi^2(1) = 4.41$, $p = .036$), as well as a Condition × Dynamic Gesture interaction for the proof ($\chi^2(1) = 10.77$, $p = .001$), insight ($\chi^2(1) = 8.85$, $p = .003$), and intuition ($\chi^2(1) = 8.96$, $p = .003$) models. Table 4 shows the regression results for the models where the interaction term was significant. Note that prior to this interaction term being added, gesture always had an overall positive, significant association with performance measures.

In model 7 in Table 4, we see that for trials where participants performed irrelevant actions (the reference category), making any gestures was not significantly associated with proof performance ($p = .70$). Likewise, for trials where participants did not gesture at all, receiving relevant actions did not have a significant association with proof performance ($p = .19$). However, in trials where participants did perform any gesture, participants who received relevant actions significantly outperformed participants who received irrelevant actions on proof performance ($d = 1.13$, 95% CI [0.43, 2.09], $p = .003$).

In model 8 in Table 4, we see that for trials where participants performed irrelevant actions (the reference category), making dynamic gestures did not have a significant association with proof performance ($p = .97$). Likewise, for trials where participants did not make any dynamic gestures, receiving relevant actions did not have a significant association with proof performance ($p = .39$). However, in trials where participants did perform dynamic gestures, participants who received relevant actions significantly outperformed participants who received irrelevant actions on proof performance ($d = 1.56$, 95% CI [0.53, 2.58], $p = .003$).

In model 9 in Table 4, we see that for trials where participants performed irrelevant actions (the reference category), making any gestures had a significant positive association with insight ($d = 0.60$, 95% CI [0.25, 0.94], $p < .001$). We also see that for trials where participants did not make any gestures, receiving relevant actions actually had a slight significantly negative effect on insight ($d = –0.48$, 95% CI [–0.15, –0.81], $p = .004$). This finding is illustrated by the fifth set of contrasting bars in Fig. 9; however, this visual suggests this may be a weak finding. In trials where participants did perform any gestures, participants who received relevant actions significantly outperformed participants who received irrelevant actions on insight ($d = 0.81$, 95% CI [0.30, 1.32], $p = .002$).

Table 4
Mixed effects logistic regression models predicting proof, insight, and intuition with gesture as a moderator

	Model 7: Proof and Any Gesture	Model 8: Proof and Dynamic Gesture	Model 9: Insight and Any Gesture	Model 10: Insight and Representational Gesture	Model 11: Insight and Dynamic Gesture	Model 12: Intuition and Dynamic Gesture
Random Effects: (St. Dev.)						
Participant ID	0.00005	0.0004	0.0004	0.00002	0.36	0.44
Conjecture	0.75	0.52	0.44	0.47	0.40	0.45
Fixed Effects:						
(Intercept)	–3.54 (0.45)	–3.33 (0.37)	–1.84 (0.25)	–1.85 (0.253)	–1.59 (0.22)	0.04 (0.21)
Irrelevant Actions	(ref.)	(ref.)	(ref.)	(ref.)	(ref.)	(ref.)
Relevant Actions	–0.67 (0.52)	–0.37 (0.43)	–0.87 (0.30)**	–0.65 (0.28)*	–0.57 (0.24)*	–0.26 (0.18)
Spatial Test Score (Centered)	0.12 (0.04)**	0.11 (0.04)**	0.11 (0.03)***	0.11 (0.03)***	0.11 (0.03)***	0.09 (0.02)***
Geometry Interest (Centered)	0.75 (0.30)*	0.81 (0.30)**	0.51 (0.17)**	0.54 (0.17)**	0.50 (0.18)**	
Phonemic Fluency (Centered)			0.10 (0.03)**	0.10 (0.03)**	0.10 (0.03)**	
Any Gesture	0.23 (0.58)		1.08 (0.32)***			
Depictive Gesture				1.38 (0.35)***		
Dynamic Gesture		0.03 (0.82)			0.41 (0.46)	–0.53 (0.42)
Relevant Actions: Any Gesture	2.28 (0.77)**		1.47 (0.47)**			
Relevant Actions: Dynamic Gesture		2.82 (0.95)**			1.84 (0.63)**	1.81 (0.63)**
Relevant Actions: Representational Gesture				1.05 (0.50)*		

*= $p < .05$.
**= $p < .01$.
***= $p < .001$.

In model 10 in Table 4, we see that for trials where participants performed irrelevant actions (the reference category), making depictive gestures had a significant positive association with insight ($d = 0.76$, 95% CI [0.38, 1.14], $p < .001$). We also see that for trials where participants did not make depictive gestures, receiving relevant actions actually had a slight significantly negative effect on insight ($d = –0.36$, 95% CI [–0.06, –0.66], $p = .020$). This finding is illustrated by the third set of contrasting bars in Fig. 9. In trials where participants did perform depictive gestures, participants who received relevant actions significantly outperformed participants who received irrelevant actions on insight ($d = 0.58$, 95% CI [0.04, 1.13], $p = .037$).

In model 11 in Table 4, we see that for trials where participants performed irrelevant actions (the reference category), making dynamic gestures did not have a significant association with insight ($p = .37$). We also see that for trials where participants did not make any dynamic gestures, receiving relevant actions actually had a slight significantly negative effect on insight ($d = –0.32$, 95% CI [–0.06, –0.57], $p = .017$). This finding is illustrated by the first set of contrasting bars in Fig. 9; however, this visual suggests this may be a weak finding. In trials where participants did perform dynamic gestures, participants who received relevant actions significantly outperformed participants who received irrelevant actions on insight ($d = 1.02$, 95% CI [0.33, 1.70], $p = .004$).

In model 12 in Table 4, we see that for trials where participants performed irrelevant actions (the reference category), making dynamic gestures did not have a significant association with intuition ($p = .21$). Likewise, for trials where participants did not make any dynamic gestures, receiving relevant actions did not have a significant association with intuition ($p = .14$). However, in trials where participants did perform dynamic gestures, participants who received relevant actions significantly outperformed participants who received irrelevant actions on intuition ($d = 1.00$, 95% CI [0.32, 1.68], $p = .004$).

Taken together, the results from models 7 to 12 are consistent with what we call the *moderated cognitive relevance hypothesis*. The moderated cognitive relevance hypothesis posits that gesture production during mathematical explanations is associated with differences in the relationship between directed actions and cognitive performance. In particular, when learners perform cognitively relevant directed actions, if they subsequently gesture, this can allow the actions to become beneficial to understanding mathematical relationships. This relationship is illustrated in Fig. 3 (bottom). Fig. 3 shows the primary difference between mediation and moderation. In the mediation model, directed actions *increase* gestures, and this increase leads to an *increase* in more valid mathematical reasoning. In the moderation model, directed actions lead directly to an increase in more valid mathematical reasoning, *but only if* gestures are present (here gestures are not necessarily *increased*).

We now in a final section of the results examine whether there is evidence that participants were aware of the connection between their relevant directed actions and the geometry conjectures.

4.4. Participant awareness of connection between conjectures and motions

We confront the question of whether participants were aware of the connection between the cognitively relevant directed motions and the conjectures they were being asked to prove, despite not being told about the connection. We examine this question in two ways. First, if participants were aware of the connection between their relevant actions and the conjecture, we might expect to see more overall gestures during the cognitively relevant trials. We did not observe this outcome. This suggests that students were not aware of the connections between the directed actions and the conjectures. We also examined each trial in the relevant condition for evidence that participants implied that their reasoning was related to the directed actions they just performed. We did not see any instances of these reports.

We had also collected data at the end of each session asking participants whether they had noticed a connection between their directed actions and the conjectures, and if so, when they noticed it, and what they thought the connection was. This was instigated by an earlier experiment (Nathan et al., 2014) that showed that, despite observing benefits of directed actions on mathematical reasoning, only a small proportion (4 out of 80) of undergraduates who performed them reported noticing any connection. None of the significant results changed when the data from that study were reanalyzed excluding those participants.

To explore this in the current study, we categorized participants' responses to the post-session questions, to see what we could observe. Note that in addition to the relevant and irrelevant directed actions used during game play, participants were instructed to make distinct arm poses to advance through the game (forward and backward) and to select among multiple-choice answers (A through D) rather than use an external keyboard or mouse, in order to maintain body calibration of the motion sensor. Among all 85 students (online Appendix C): 19 articulated *only* a connection between their motions and a conjecture for which they had received cognitively *relevant* motions; 16 articulated a connection *only* to a conjecture for which they had received *irrelevant* motions and/or to *other arm movements* that were used to control the game (such as advancing screens and responding to multiple choice prompts); and 12 made such connections for both relevant and irrelevant motions. Drilling down a bit, we found that even for the instances where students cited the cognitively relevant directed actions as being relevant to the mathematical conjecture, there was little evidence that any student explicitly noticed the *mathematical relationship* the motions were intended to embody (e.g., similarity, relationship between angle and side). Most of the relevant connections were general observations that they had made a triangle with their arms and gotten a triangle conjecture, or that they had made a parallelogram with their arms and gotten a parallelogram conjecture. These comments were similar to the observations of students who made connections to irrelevant motions and game control motions, suggesting that it was unlikely they noticed any relational connections. The fact that students performed both irrelevant and relevant motions interleaved made explicitly catching onto the connection difficult for them.

To be thorough, though, we removed the 19 students who we termed as "possible noticers" of the connection between the cognitively relevant actions and the conjectures from the dataset. We then reran all analyses with this reduced dataset. We found that results were the same, except for one interaction (condition by depictive gesture interaction for the insight outcome) that changed in its p-value from .036 to .054—which is not surprising given the reduction in power. These findings again suggest that the results of this study are not being driven by students being explicitly aware of the connection between the conjectures and the cognitively relevant directed motions. We thus close by considering the theoretical and educational implications of these findings.

5. Discussion and significance

The present study explored three related hypotheses regarding the relation of action and thought. The *direct cognitive relevance hypothesis* posited that the cognitive relevance of actions has a direct effect on performance. The *mediator hypothesis* proposed that the benefits of cognitively relevant actions would lead to more gestures during explanations than irrelevant actions, which would in turn lead to higher performance. The *moderator hypothesis* proposed that cognitively relevant directed actions would only lead to superior reasoning when gestures were present during students' explanations. The findings of this within-subjects experiment showed that cognitive relevance indeed matters and depends on the moderating role of explanatory gestures for the effect of relevant directed actions to deliver on their benefits. No such benefit was evident for cognitively irrelevant actions, even though they were produced by rearranging the sequences of cognitively relevant actions, and, therefore, were of comparable motor complexity. The pattern of results does not seem to be due to students' awareness of the mathematical connection between the conjectures and the cognitively relevant directed motions.

Our results suggest that cognitively relevant directed actions did not directly *cause* learners to gesture more during subsequent explanations or cause them to engage in superior mathematical reasoning when giving proofs during subsequent explanations. Rather, directed actions may be necessary for students to better schematize the invariant mathematical relations embodied by the directed actions using gestures during explanations in order for the cognitive relevance of earlier actions to influence mathematical cognition. In this way, gestures that "replay" the directed actions in some manner may serve as a bridge between concrete directed actions and the generalized reasoning required for geometric proof, as we describe further in the next section. This effect was most broadly consistent across the range of outcome measures—intuition, insight, and proof production—when students produced dynamic gestures during reasoning, suggesting the schematization of simulating mathematical transformations through dynamic gesture may be particularly useful in the domain of geometric proofs. However, this is not the only possible explanation for our pattern of findings—as we discuss later, gestures may be an index of a highly related but slightly different moderator, that is, action-based mental simulations.

We also found that irrelevant actions were in some cases associated with slightly stronger performance on insight than relevant actions when participants did not gesture. This finding was somewhat weak and may be driven by the near-floor performance of participants when they did not gesture; logistic models can magnify even small differences when they appear near the tails of the distribution. However, it could suggest that the relevant actions caused some sort of interference for those who did not engage in gestural replays, although speculating on why or how is beyond the scope of the kinds of data we collected.

5.1. Gesture as a moderator of the effectiveness of cognitively relevant motions

An important question is *why* the presence of gestures moderated the effectiveness of directed actions. The presence of explanatory gestures seemed to act as a *moderator* for the effects of directed actions, in that directed actions only show differential effects on reasoning when opportunities for explanatory gestures are present. Gestures are special sorts of actions that convey meaning rather that operating directly on material objects in the world. Gestures can support reasoning by focusing one's limited cognitive resources on task relevant information gleaned from effective actions, which can facilitate the construction of schemas used in future task performance. In other words, gestures can highlight important information that make actions effective and enable action schemas to generalize. In this way, gestures may serve a moderating role on the effect that actions have on cognition, by schematizing the most important qualities of task-relevant actions—either physical actions or simulated actions—and that can then benefit task-relevant reasoning processes. Thus, the information provided by gesture production during explanations can change the effects of directed actions on subsequent reasoning.

An example of a moderation effect for gestures is in Beilock and Goldin-Meadow (2010), who directed participants to solve weighted versions of the 4-disk Tower of Hanoi task across two sessions. In the first session, the diameters of the disks correlated with their weight. The smallest was the lightest and could be moved with one hand, while the largest was the heaviest and required two hands to lift and move. The disks were identical in the second session for the no-switch group, but the relationship of the weights to size was reversed for the switch group, with lifting the smallest disk now requiring two hands. Some participants were prompted to explain their solutions after the first session and prior to the second session, and those explanations often included one-handed gestures for moving the smallest disk and two-handed gestures for moving the larger disk. The investigators observed no performance differences between the two problem-solving sessions for the no-switch group as well as no differences for participants who were not prompted to explain their solutions. However, those in the switch group who generated gesture-rich explanations showed significant performance decline during the second session. Furthermore, the more they gestured with one-handed moves for the small disk during their explanations of their session 1 solutions, the worse they did in session 2. It seemed that engaging in these "gestural replays" of their actions during explanations—even when those replays were not identical recreations of the original actions—changed participants' encoding for the mathematical principles of the task and subsequent task performance. Thus, gestures can influence one's future reasoning by bringing perceptual-motor information obtained from actions into one's cognitive account. Gestures "provide a bridge between concrete actions and more abstract representation" (Goldin-Meadow & Beilock, 2010, p. 672).

This may suggest that when learners can make or effectively learn from "gesture replays" of previous cognitively relevant actions during their explanations, their subsequent cognitive performance can be enhanced. Gestural replays are hypothesized to be the primary mechanism for carrying out the embodied simulations that support geometric insight. Cognitively relevant directed actions may allow learners to *more effectively utilize their explanatory gestures during gestural replays* to engage deeply with the ideas of the conjecture.

The way we conceptualize gestural replays, they do not necessarily need to precisely resemble the directed actions. For example, in prior work, learners with an embodied understanding of two gears turning against each other, pushing in opposite directions, may gesture their understanding of the mathematical principle of parity in a way that does not involve turning gears at all—they may simply tap back and forth with their index finger to show alternation (see Nathan et al., 2014). Gestural replays can undergo levels of refinement from the original directed actions. However, we saw gestural replays that overtly resembled the *relevant* directed action sequences in approximately one-third of our cases. We did not see the same trend for *irrelevant* directed actions; participants may have been unlikely to replay irrelevant directed actions because they are not commensurate with the movements that aligned with the specific conjectures. Our impressions are that students were primarily focused on making meaning of the conjectures, rather than inferring the meaning of the directed actions, as detailed in Section 4.5. In contrast, the relevant directed actions they performed did align with the mathematical ideas posed by each conjecture.

5.2. *Emerging theories of grounded and embodied cognition*

The findings contribute to emerging theories of grounded and embodied cognition in two ways. First, grounded and embodied cognition theories need to explicate the nature of the cognitive relevance of actions. Second, there is a need to understand the differential effects of actions and gestures on cognition. Both issues rest on a deepening understanding of the effects specific body movements have on thinking, and the ways that motoric processes shape reasoning and learning.

Kita et al. (2017) proposed the gesture-for-conceptualization hypothesis as an account for how representational gestures—such as the dynamic gestures we observed—serve the speaker. They provide evidence that a speaker's gestures *schematize* information generated via task-relevant actions. In reviewing the study conducted by Beilock and Goldin-Meadow (2010) on the use of weighted Tower of Hanoi discs, Kita and colleagues (2017, p. 256) concluded:

> that gesturing about actions exerted a stronger influence on how action-relevant information was mentally represented than actually performing the actions. Put another way, weight information was incorporated into the schematized spatio-motor representations (one-handed vs. two-handed movement) that participants constructed in the gesture condition, so the shift in weight was more problematic for them.

Separately, Nathan (2017) proposed ACT as an account that relies on the reciprocal exchange of energy and information flowing between motoric and reasoning processes that are activated during the formation of goal-directed actions. Transduction is ubiquitous in physical systems, such as the dual relationship between motors and generators, and physiological systems, such as the influence facial muscle paralysis from Botox injections has on cognitive processing of text with emotional information (Havas, Glenberg, Gutowski, Lucarelli, & Davidson, 2010). ACT describes how actions performed on either real or imagined entities generate feedforward (predictive) and feedback (responsive) signals that can activate cognitive states that are associated with important relationships and behaviors of the entities, such as model-based reasoning and inference-making (Nathan & Martinez, 2015). As such, the architecture of embodied cognition that follows from ACT is one that positions mental processes that are traditionally treated as "cognitive" on the same footing as the processes for regulating action that have traditionally been treated as "sensorimotor" exclusive of cognition. As described by Nathan (2021, p. 103) "In a transduction model of cognition and action, neither the mind nor the action system occupies a central processing role, or, more aptly, they both do, as each serves a role shaping the state of the other."

In light of past and current findings, we propose a model of the processes by which actions and explanatory gestures influence reasoning, integrating the feedforward account from ACT with the schematization processes of gesture-for-conceptualization from Kita et al. (2017). This model draws on the design of the HMOSAIC system (Haruno, Wolpert, & Kawato, 2003), an architecture for regulating motor control in service of goal directed behavior in a changing and uncertain environment. HMOSAIC adopts a predictive stance toward control in order to both monitor one's actions in the world and anticipate their likely effects on the body and the environment. The HMOSAIC architecture has inspired motor control-based accounts for a range of complex behaviors, including action production in an uncertain environment, social interaction, and Theory of Mind (Wolpert, Doya, & Kawato, 2003), and action-based models of language comprehension (Glenberg & Gallese, 2012).

We build off of this motor control architecture to initially describe the processes that influence one's cognitive state when a player of our embodied video game, *The Hidden Village*, pursues a goal to mimic the directed actions of the in-game avatars. We then follow with a process account for how one's cognitive state is influenced by the goal-directed actions generated during one's explanatory gesture production.

As shown in Fig. 11, the process of producing a directed action starts with a goal structure, a cognitive state with the intention of changing the physical state of the world, along with a set of action plans to achieve that goal, subject to the preconditions of the current context and the actor involved. Each action plan forms its own motor program. The case of mimicking directed actions is illustrated as instances of specific actions, such as raising one's forearms with elbows bent (flexion) so hands are positioned just above the shoulders, (step 1) and pivoting laterally at the elbow (step 2). Performing the actions generates an efferent copy that transmits the information about the actual movements performed.

The reality of how an action is carried out, however, is imprecise, subject to factors of both the agent and the environment and must garner continual monitoring for its success. Waiting to determine if the outcome is achieved before making corrections is nonoptimal, since reacting to feedback (e.g., through proprioception and vision) is slow and this time lag might put the agent at risk (e.g., being hit by an object). As the architects of the HMOSAIC system have realized, the problem of response lag from reactive feedback can be addressed by proactively using feedforward. Feedforward mechanisms are commonly found in the motor system and serve movement prediction, monitoring, and control until each subgoal of the action plan is achieved. To achieve this proactive orientation alongside the reactive orientation achieved through monitoring, each action plan in HMOSAIC includes the generation of multiple,

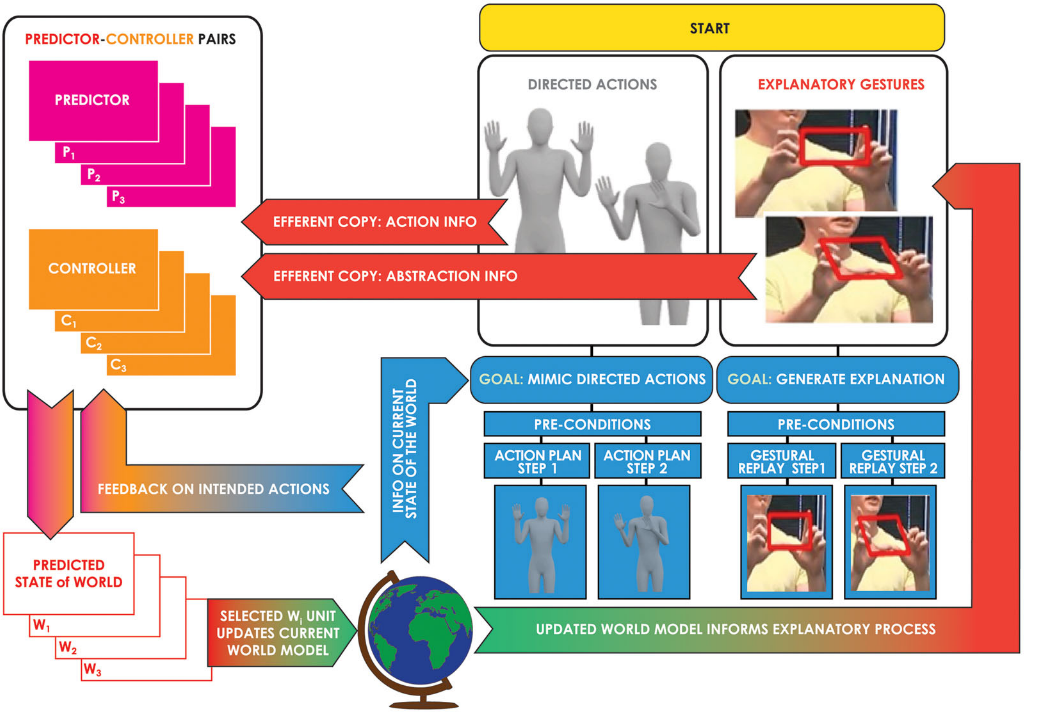


Fig. 11. An illustration of how directed actions (middle column) and explanatory gestures (right-most column) influence thought and update an agent's world model.

paired predictor-controller modules. Each predictor-controller module (i.e., module$_i$) is a small motor program that both guides (i.e., controls, via controller c_i) the particulars of the motor behavior (e.g., solving the reverse kinematics equations for each of the joints along the arm and hand to match a specific directed action) and anticipates (i.e., predicts, via predictor p_i) the set of highly likely next states of the motor system. Each p_i-c_i module provides feedforward and proprioceptive feedback signals that regulate muscle behavior in order to ensure progress toward the intended goal. Notably, the various predictor-controller modules are in continuous competition with each other. The system rewards those p_i-c_i modules that are most successful at guiding the motor system within the dynamic environment in real time.

Each p_i-c_i module provides, in essence, a simulation of the many plausible future states of the world, W_i that are compatible with outcomes conforming to the current action plan. Together, the array of currently active predictor–controller pairs form a set of plausible inferences about the world. When a predictor–controller pair converges with the currently active goal, that goal-directed action plan concludes, and the participant's cognitive state for their current world model (depicted by the globe) is updated, replaced with the W_i state that has been most highly activated.

The process account for the moderator role of explanatory gestures utilizes the same basic architecture (Fig. 11). As a moderator, the production of explanatory gestures interacts with the ways that directed actions influence participants' mathematical reasoning. Our proposed model draws on the schematizing function of gestures that serve to highlight certain properties of actions and concepts they simulate. In this case, the goal structure for generating an explanatory gesture as part of an explanation generates a set of action plans that include gestural replays. As an action that is to be faithfully executed, producing a gestural replay generates an efferent copy that elicits a corresponding set of predictor-controller modules. However, unlike the modules generated for directed actions, these gestures are not intended to match a specific movement sequence. Rather, they can convey any of a broad set of movements that enact *idealized forms* or abstractions that are congruent with the semantic and spatial information of the explanation that require the least effort for someone to perceive the invariant properties of interest (Fyfe & Nathan, 2019). The schematizing function of gestural replays naturally highlights idealized and invariant qualities of the mathematical objects and relations they are meant to convey. For example, as shown in Fig. 11, directed actions cuing arm movements may be gesturally replayed with other body parts (in this case, fingers) that preserve some core invariant properties of the area of the entire class of parallelograms.

As with the sequence described earlier, generation of predictor-controller modules produces a simulation of the future states of the world model, with competition between the modules leading to the most plausible update of the participant's world model. In general, relevant directed actions tend to lead to an updated world model that is more accurate in terms of how things will actually change from one's actions. Engaging the explanation process when operating with a more accurate world model means that the gestural replays that are produced are more likely to highlight the most relevant invariant information. This, in turn, increases the likelihood that a participant will generate a mathematically valid proof that conforms to the scoring criteria of being logical, operational, and generalizable.

One additional theoretical consideration is the source of cognitively relevant directed actions. As noted, the directed actions implemented in the game were inspired by the gestures that successful problem-solvers produced when proving these conjectures in prior investigations. We hypothesized that these would be helpful for future students, and indeed, this was what we found. Why these movements are effective likely has to do with the enactive processes of those original participants who were engaged in these proof activities. By one account (e.g., Abrahamson & Sánchez-García, 2016), these earlier participants solved a motor coordination problem oriented toward their perceived affordances of the imagined objects, which served as "attentional anchors" (p. 216) or conscious constraints on the multitude of ways those movements could be performed. Future research on the coordination dynamics of these movement schemes may reveal important insights into these more basic embodied processes for mathematical reasoning.

5.3. *Implications for educational interventions*

Empirical support for the nature of embodied thinking and learning provides guidance for the design and development of future learning interventions (Goldstone, Landy, & Son, 2008). In particular, explicit prompting for students to turn directed actions into personally meaningful hand gestures may facilitate the effectiveness of interventions that direct motion, as was done in previous classroom-based interventions with *The Hidden Village* (Kirankumar et al., 2021; Walkington et al., 2021). In addition, students' individual characteristics like gesture threshold (Hostetter & Alibali, 2008, 2019) might be critical to understanding the differential impact of embodied learning interventions on student learning outcomes.

It is difficult to untangle from this study whether the gesture production during explanation was itself the causal mechanism that allowed directed actions to be effective, or if it implicated some highly related action-oriented factor. Gesture usage might index whether students are thinking about the conjecture in a transformative, action-based way instead of in a procedural or algorithmic manner, and this tendency itself may be what allows directed actions to gain their benefit. One might imagine that for students who are thinking of geometric conjectures in a rote, definitional manner without engaging in mental simulations on objects, performing directed arm motions may not be particularly helpful—the associated perceptuo-motor resources might go unused. However, students who are performing mental simulations of actions, which can in turn give rise to gestures, may be more likely to benefit from directed actions. It seems less likely that gesture is a proxy for some deep form of mathematical understanding or engagement. We found only a weak relationship between gesture and valid mathematical reasoning for students who performed irrelevant actions, suggesting this is not the case. Additionally, we did not see more gestures in the relevant action condition, which suggests that gestures representing deeper understanding or engagement spurred by relevant actions is unlikely. And finally, in our models, we controlled or tested the impact of controlling for a variety of student mathematical understanding covariates, including spatial score, geometry pretest, geometry interest (which is strongly related to engagement), and other math background variables.

It is notable that this study achieved significant moderated effects from directed actions on tasks in a relatively complex mathematical domain. Directed motion seems to be a fruitful area for future research, especially if its benefits can be realized for all learners.

The present study was conducted with individual students in order to be able to tightly control and understand individual learning processes. In other research, we found that movement-based game play with partners dramatically increases the rates of gesture during explanations (Abrahamson et al., 2020; Walkington et al., 2019), and the length and complexity of those explanations, even though only one learner typically performs the prior directed actions. The ways in which directed actions, gesture, and performance relate may be fundamentally different in a collaborative context. There are also important differences in observing versus performing actions and gestures in these contexts (Alibali & Nathan, 2018).

This work suggests the potential benefits of embodied learning interventions that use cognitively relevant actions and prompt learners to generate explanatory gestures. Since gesture rates often increase in a social context, there may be additional benefits incurred by engaging students to collaborate. This can also lead to the generation of interpersonally coordinated collaborative gestures, which can also facilitate mathematical reasoning (Abrahamson et al., 2020). Another approach is to engage groups of students as collaborative codesigners of cognitively relevant directed actions, as they create new game content with gestures for their peers (Kirankumar et al., 2021; Walkington et al., 2021). In this approach, students engage in rich discussions of the cognitive relevance of the actions that would help other students apprehend their intended mathematical ideas. This encourages a form of distributed transfer from one group of students to another, mediated by active engagement with directed actions that are cognitively relevant to the mathematics at hand. A number of emerging theoretical accounts of transfer are based on embodiment and collaborative embodiment (e.g., Goldstone et al., 2008; Walkington et al., 2019; Nathan et al., 2021).

Investigations of this sort can contribute to both emerging theories of mathematical reasoning as embodied processes, and to designing a class of embodied interventions to foster thinking and learning. As research on embodied learning matures, increasingly complex, situated, collaborative, and conceptually rich interventions will be developed for key ideas to be taught in classrooms and the workplace. For each of these contributions, the cognitive relevance of the actions performed are likely to be of importance.

Acknowledgments

The research reported here was supported by the Institute of Education Sciences, U.S. Department of Education, through grant R305A160020 to University of Wisconsin – Madison. The opinions expressed are those of the authors and do not represent views of the Institute or the U.S. Department of Education. We would like to thank Peter Steiner, Michael Swart, Dawn Woods, Diana Vu, and Kasi Holcomb-Webb for their contributions to this research, as well as the members of our advisory board.

References

Abrahamson, D. (2004). Embodied spatial articulation: A gesture perspective on student negotiation between kinesthetic schemas and epistemic forms in learning mathematics. In D. E. McDougall & J. A. Ross (Eds.), *Proceedings of the 26th Annual Meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education* (pp. 791–797). Toronto, ON: Preney.

Abrahamson, D., & Bakker, A. (2016). Making sense of movement in embodied design for mathematics learning. *Cognitive Research: Principles and Implications*, *1*(1), 1–13.

Abrahamson, D., Nathan, M. J., Williams–Pierce, C., Walkington, C., Ottmar, E. R., Soto, H., & Alibali, M. W. (2020). The future of embodied design for mathematics teaching and learning. *Frontiers in Education*, *5*, 147.

Abrahamson, D., & Sánchez-García, R. (2016). Learning is moving in new ways: The ecological dynamics of mathematics education. *Journal of the Learning Sciences*, *25*(2), 203–239.

Adams, A. M., Glenberg, A. M., & Restrepo, M. A. (2019). Embodied reading in a transparent orthography. *Learning and Instruction*, *62*, 27–36.

Alibali, M. W., Kita, S., & Young, A. J. (2000). Gesture and the process of speech production: We think, therefore we gesture. *Language and Cognitive Processes*, *15*(6), 593–613.

Alibali, M. W., & Nathan, M. J. (2018). Embodied cognition in learning and teaching: Action, observation, and imagination. In F. Fischer, S. Goldman, C. Hmelo-Silver, & P. Riemann (Eds.), *International handbook of the learning sciences* (pp. 75–85). New York: Routledge/Taylor & Francis.

Baron, R. M., & Kenny, D. A. (1986). The moderator–mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology*, *51*, 1173–1182.

Barrett, T. J., Stull, A. T., Hsu, T. M., & Hegarty, M. (2015). Constrained interactivity for relating multiple representations in science: When virtual is better than real. *Computers & Education*, *81*, 69–81.

Barsalou, L. W. (1999). Perceptual symbol systems. *Behavioral and Brain Sciences*, *22*(4), 577–660.

Barsalou, L. W. (2008). Grounded cognition. *Annual Review of Psychology*, *59*, 617–645.

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2014). lme4: Linear mixed-effects models using Eigen and S4. R package version 1.1-7. http://CRAN.R-project.org/package=lme4.

Beilock, S. L., & Goldin-Meadow, S. (2010). Gesture changes thought by grounding it in action. *Psychological Science*, *21*(11), 1605–1610.

Berch, D. B., Foley, E. J., Hill, R. J., & Ryan, P. M. (1999). Extracting parity and magnitude from Arabic numerals: Developmental changes in number processing and mental representation. *Journal of Experimental Child Psychology*, *74*(4), 286–308.

Bruderer, A. G., Danielson, D. K., Kandhadai, P., & Werker, J. F. (2015). Sensorimotor influences on speech perception in infancy. *Proceedings of the National Academy of Sciences*, *112*(44), 13531–13536.

Casasanto, D., & Chrysikou, E. G. (2011). When left is "right" motor fluency shapes abstract concepts. *Psychological Science*, *22*(4), 419–422.

Casasanto, D., & de Bruin, A. (2019). Metaphors we learn by: Directed motor action improves word learning. *Cognition*, *182*, 177–183.

Casasanto, D., & Henetz, T. (2012). Handedness shapes children's abstract concepts. *Cognitive Science*, *36*(2), 359–372.

Chinn, S. (2000). A simple method for converting an odds ratio to effect size for use in meta-analysis. *Statistics in Medicine*, *19*(22), 3127–3131.

Church, R. B., & Goldin-Meadow, S. (1986). The mismatch between gesture and speech as an index of transitional knowledge. *Cognition*, *23*(1), 43–71.

Clements, D. H., & Battista, M. T. (1992). Geometry and spatial reasoning. D. A. Grouws (Ed.) *Handbook of research on mathematics teaching and learning* (pp. 420–464).

Cook, S. W., & Goldin-Meadow, S. (2006). The role of gesture in learning: Do children use their hands to change their minds? *Journal of Cognition and Development*, *7*(2), 211–232.

Cook, S. W., & Tanenhaus, M. K. (2009). Embodied communication: Speakers' gestures affect listeners' actions. *Cognition*, *113*(1), 98–104.

Cutica, I., & Bucciarelli, M. (2008). The deep versus the shallow: Effects of co-speech gestures in learning from discourse. *Cognitive Science*, *32*, 921–935.

desRosiers, G., & Kavanagh, D. (1987). Cognitive assessment in closed head injury: Stability, validity and parallel forms for two neuropsychological measures of recovery. *International Journal of Clinical Neuropsychology*, *9*, 162–173.

Donovan, A., & Alibali, M. W. (2018). Do actions during math learning leave a legacy in gesture?. In T. T. Rogers, M. Rau, X. Zhu, & C. W. Kalish (Eds.), *Proceedings of the 40th Annual Conference of the Cognitive Science Society* (pp. 2848). Madison, WI: Cognitive Science Society.

Ekstrom, R. B., French, J. W., Harman, H., & Derman, D. (1976). *Kit of factor-referenced cognitive tests* (rev. ed.). Princeton, NJ: Educational Testing Service.

Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods*, *41*, 1149–1160.

Fayol, M., Barrouillet, P., & Marinthe, C. (1998). Predicting arithmetical achievement from neuro-psychological performance: A longitudinal study. *Cognition*, *68*(2), B63–B70.

Fritz, M. S., & MacKinnon, D. P. (2007). Required sample size to detect the mediated effect. *Psychological Science*, *18*(3), 233–239.

Fyfe, E. R., & Nathan, M. J. (2019). Making "concreteness fading" more concrete as a theory of instruction for promoting transfer. *Educational Review*, *71*(4), 403–422.

Garcia, N., & Infante, N. E. (2012). Gestures as facilitators to proficient mental modelers. In L. R. Van Zoest, J.-J. Lo, & J. L. Kratky (Eds.), *Proceedings of the 34th Annual Meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education* (pp. 289–295).

Glenberg, A., Willford, J., Gibson, B., Goldberg, A., & Zhu, X. (2012). Improving reading to improve math. *Scientific Studies of Reading*, *16*(4), 316–340.

Glenberg, A. M. (1997). What memory is for. *Behavioral and Brain Sciences*, *20*(1), 1–19.

Glenberg, A. M., & Gallese, V. (2012). Action-based language: A theory of language acquisition, comprehension, and production. *Cortex*, *48*(7), 905–922.

Glenberg, A. M., Gutierrez, T., Levin, J. R., Japuntich, S., & Kaschak, M. P. (2004). Activity and imagined activity can enhance young children's reading comprehension. *Journal of Educational Psychology*, *96*(3), 424.

Glenberg, A. M., & Robertson, D. A. (1999). Indexical understanding of instructions. *Discourse Processes*, *28*(1), 1–26.

Göksun, T., Goldin-Meadow, S., Newcombe, N., & Shipley, T. (2013). Individual differences in mental rotation: What does gesture tell us? *Cognitive Processing*, *14*(2), 153–162.

Goldin-Meadow, S. (2003). *Hearing gesture: How our hands help us think*. Harvard University Press.

Goldin-Meadow, S., & Beilock, S. L. (2010). Action's influence on thought: The case of gesture. *Perspectives on Psychological Science*, *5*(6), 664–674.

Goldin-Meadow, S., Cook, S. W., & Mitchell, Z. A. (2009). Gesturing gives children new ideas about math. *Psychological Science*, *20*(3), 267–272.

Goldstone, R. L., Landy, D., & Son, J. Y. (2008). A well-grounded education: The role of perception in science and mathematics. In M. de Vega, A. Glenberg, & A. Graesser (Eds.), *Symbols, embodiment and meaning* (pp. 327–355).

Harel, G., & Sowder, L. (1998). Students' proof schemes. In E. Dubinsky, A. Schoenfeld, & J. Kaput (Eds.), *Research on collegiate mathematics education* (Vol. *III*, pp. 234–283). Providence, RI: American Mathematical Society.

Harel, G., & Sowder, L. (2007). Toward comprehensive perspectives on the learning and teaching of proof. In F. Lester (Ed.), *Second handbook of research on mathematics teaching and learning*. Reston, VA: National Council of Teachers of Mathematics. 805–842.

Haruno, M., Wolpert, D. M., & Kawato, M. (2003). Hierarchical MOSAIC for movement generation. In *International Congress Series* (Vol. *1250*, pp. 575–590). Elsevier.

Havas, D. A., Glenberg, A. M., Gutowski, K. A., Lucarelli, M. J., & Davidson, R. J. (2010). Cosmetic use of botulinum toxin-A affects processing of emotional language. *Psychological Science*, *21*(7), 895–900.

Hidi, S., & Renninger, K. A. (2006). The four-phase model of interest development. *Educational Psychologist*, *41*(2), 111–127.

Hollebrands, K. F. (2007). The role of a dynamic software program for geometry in the strategies high school mathematics students employ. *Journal for Research in Mathematics Education*, *38*(2), 164–192.

Hostetter, A. B., & Alibali, M. W. (2007). Raise your hand if you're spatial: Relations between verbal and spatial skills and gesture production. *Gesture*, *7*(1), 73–95.

Hostetter, A. B., & Alibali, M. W. (2008). Visible embodiment: Gestures as simulated action. *Psychonomic Bulletin & Review*, *15*(3), 495–514.

Hostetter, A. B., & Alibali, M. W. (2019). Gesture as simulated action: Revisiting the framework. *Psychonomic Bulletin & Review*, *26*(3), 721–752.

Jones, K., & Tzekaki, M. (2016). Research on the teaching and learning of geometry. In A. Gutiérrez, J. Leder, & P. Boero (Eds.), *The second handbook of research on the psychology of mathematics education* (pp. 109–149). Brill Sense.

Kenny, D. A., Kashy, D. A., & Bolger, N. (1998). Data analysis in social psychology. In D. Gilbert, S. Fiske, & G. Lindzey (Eds.), *The handbook of social psychology* (Vol. *1*, pp. 233–265). Boston, MA: McGraw-Hill.

Kim, M., Roth, W. M., & Thom, J. (2011). Children's gestures and the embodied knowledge of geometry. *International Journal of Science and Mathematics Education*, *9*(1), 207–238.

Kim, S., Jiang, Y., & Song, J. (2015). The effects of interest and utility value on mathematics engagement and achievement. In K. A. Renninger, M. Nieswandt, & S. Hidi (Eds.), *Interest in mathematics and science learning* (pp. 63–78). Washington, DC: American Educational Research Association.

Kirankumar, V., Sung, H., Swart, M., Kim, D., Xia, F., Kwon, O. H., … Walkington, C. (2021). Embodied transmission of ideas: Collaborative construction of geometry content and mathematical thinking. In C. E. Hmelo-Silver, B. De Wever, & J. Oshima (Eds.), *Proceedings of the 14th International Conference on Computer-Supported Collaborative Learning - CSCL 2021* (pp. 177–180). Bochum: International Society of the Learning Sciences.

Kita, S., Alibali, M. W., & Chu, M. (2017). How do gestures influence thinking and speaking? The gesture-for-conceptualization hypothesis. *Psychological Review*, *124*(3), 245.

Lakoff, G., & Núñez, R. E. (2000). *Where mathematics comes from: How the embodied mind brings mathematics into being*. New York: Basic Books.

Linnenbrink-Garcia, L., Durik, A., Conley, A., Barron, K., Tauer, J., Karabenick, S., & Harackiewicz, J. (2010). Measuring situational interest in academic domains. *Educational Psychological Measurement*, *70*, 647–671.

Marghetis, T., Edwards, L. D., & Núñez, R. (2014). More than mere handwaving. edited by Laurie D. Edwards, Deborah Moore-Russo, Francesca Ferrara, *Emerging perspectives on gesture and embodiment in mathematics* (pp. 227–246).

Marghetis, T., & Núñez, R. (2013). The motion behind the symbols: A vital role for dynamism in the conceptualization of limits and continuity in expert mathematics. *Topics in Cognitive Science*, *5*(2), 299–316.

McNeill, D. (1992). *Hand and mind: What gestures reveal about thought*. Chicago, IL: University of Chicago Press.

Menary, R. (2010). Introduction to the special issue on 4E cognition. *Phenomenology and the Cognitive Sciences*, *9*(4), 459–463.

Nathan, M. J. (2017). One function of gesture is to make new ideas: Evidence for reciprocity between action and cognition. In R. B. Church, M. W. Alibali, & S. D. Kelly (Eds.), *Why gesture? How the hands function in speaking, thinking and communicating* (Vol. *8*, pp. 175–196). Philadelphia, PA: John Benjamins Publishing Company.

Nathan, M. J. (2021). *Foundations of embodied learning: A paradigm for education*. Routledge.

Nathan, M. J., Alibali, M. W., & Church, R. B. (2017). Making and breaking common ground: How teachers use gesture to foster learning in the classroom. In R. B. Church, M. W. Alibali, & S. D. Kelly (Eds.), *Why gesture? How the hands function in speaking, thinking and communicating* (pp. 285–316). Philadelphia, PA: John Benjamins Publishing Company.

Nathan, M. J., & Martinez, C. V. (2015). Gesture as model enactment: The role of gesture in mental model construction and inference making when learning from text. *Learning: Research and Practice*, *1*(1), 4–37.

Nathan, M. J., Schenck, K., Vinsonhaler, R., Michaelis, J., Swart, M., & Walkington, C. (2020). Embodied geometric reasoning: Dynamic gestures during intuition, insight, and proof. *Journal of Educational Psychology*, *113*(5), 929–948.

Nathan, M. J., Schenck, K. E., Vinsonhaler, R., Michaelis, J. E., Swart, M. I., & Walkington, C. (2021). Embodied geometric reasoning: Dynamic gestures during intuition, insight, and proof. *Journal of Educational Psychology*, *113*(5), 929.

Nathan, M. J., Walkington, C., Boncoddo, R., Pier, E. L., Williams, C. C., & Alibali, M. W. (2014). Actions speak louder with words: The roles of action and pedagogical language for grounding mathematical proof. *Learning and Instruction*, *33*, 182–193.

Nemirovsky, R., Ferrara, F., Ferrari, G., & Adamuz-Povedano, N. (2020). Body motion, early algebra, and the colours of abstraction. *Educational Studies in Mathematics*, *104*, 261–283.

Newen, A., De Bruin, L., & Gallagher, S. (2018). *The Oxford handbook of 4E cognition*. Oxford University Press.

Ng, O., & Sinclair, N. (2015a). Young children reasoning about symmetry in a dynamic geometry environment. *ZDM Mathematics Education*, *47*(3), 421–434.

Novack, M. A., Congdon, E. L., Hemani-Lopez, N., & Goldin-Meadow, S. (2014). From action to abstraction: Using the hands to learn math. *Psychological Science*, *25*(4), 903–910.

Petrick, C. J. (2012). *Every body move: Learning mathematics through embodied actions* [Doctoral dissertation]. University of Texas at Austin.

Pier, E. L., Walkington, C., Clinton, V., Boncoddo, R., Williams–Pierce, C., Alibali, M. W., & Nathan, M. J. (2019). Embodied truths: How dynamic gestures and speech contribute to mathematical proof practices. *Contemporary Educational Psychology*, *58*, 44–57.

Pulvermüller, F. (2005). Brain mechanisms linking language and action. *Nature Reviews Neuroscience*, *6*(7), 576–582.

R Core Team. (2018). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.

Rav, Y. (1999). Why do we prove theorems? *Philosophia Mathematica*, *7*(1), 5–41.

Reeve, R., & Humberstone, J. (2011). Five- to 7-year-olds' finger gnosia and calculation abilities. *Frontiers in Psychology*, *2*, 359.

Reis, H. M., Borges, S. S., Durelli, V. H., Moro, L. F. D. S., Brandao, A. A., Barbosa, E. F., & Bittencourt, I. I. (2012). Towards reducing cognitive load and enhancing usability through a reduced graphical user interface for a dynamic geometry system: An experimental study. In 2012 IEEE International Symposium on Multimedia (pp. 445–450). IEEE.

Renninger, K. A., & Hidi, S. E. (2015). *The power of interest for motivation and engagement*. Routledge.

Sarama, J., & Clements, D. H. (2009). "Concrete" computer manipulatives in mathematics education. *Child Development Perspectives*, *3*(3), 145–150.

Shapiro, L. (2019). *Embodied cognition*. Routledge.

Sinclair, N., Bartolini Bussi, M. G. B., de Villiers, M., Jones, K., Kortenkamp, U., Leung, A., & Owens, K. (2016). Recent research on geometry education: An ICME-13 survey team report. *ZDM Mathematics Education*, *48*(5), 691–719.

Smith, L. B. (2005). Action alters shape categories. *Cognitive Science*, *29*(4), 665–679.

Snijders, T., & Bosker, R. (1999). *Multilevel analysis*. Sage.

Snijders, T. A. B., & Bosker, R. J. (1999). *Multilevel analysis: An introduction to basic and advanced multilevel modeling*. Thousand Oaks, CA: Sage Publications.

Thomas, L. E., & Lleras, A. (2007). Moving eyes and moving thought: On the spatial compatibility between eye movements and cognition. *Psychonomic Bulletin & Review*, *14*(4), 663–668.

Thomas, L. E., & Lleras, A. (2009). Swinging into thought: Directed movement guides insight in problem solving. *Psychonomic Bulletin & Review*, *16*(4), 719–723.

Toomarian, E. Y., Meng, R., & Hubbard, E. M. (2019). Individual differences in implicit and explicit spatial processing of fractions. *Frontiers in Psychology*, *10*, 596.

Walkington, C., Chelule, G., Woods, D., & Nathan, M. J. (2019). Collaborative gesture as a case of extended mathematical cognition. *Journal of Mathematical Behavior*, *55*, 1–20.

Walkington, C., Nathan, M. J., Holcomb-Webb, K., Wang, M., & Kwon, O. (2021). Collaborative creation of embodied geometry proofs for a motion capture game. Paper presented at the 2021 Annual Meeting of the American Educational Research Association (AERA).

Walkington, C., Woods, D., Nathan, M. J., Chelule, G., & Wang, M. (2019). Does restricting hand gestures impair mathematical reasoning? *Learning and Instruction*, *64*, 1–16 https://doi.org/10.1016/j.learninstruc.2019.101225

Walkington, C. A., Nathan, M. J., Wolfgram, M., Alibali, M. W., & Srisurichan, R. (2014). Bridges and barriers to constructing conceptual cohesion across modalities and temporalities: Challenges of STEM integration in the precollege engineering classroom. In J. Strobel, S. Purzer, & M. Cardella (Eds.), *Engineering in pre-college settings: Research into practice* (pp. 183–209). West Lafayette, IN: Purdue University Press.

Williams-Pierce, C. C., Pier, E. L., Walkington, C., Boncoddo, R., Clinton, V., Alibali, M. W., & Nathan, M. J. (2017). What we say and how we do: Action, gesture, and language in proving. *Journal for Research in Mathematics Education*, *48*(3), 248–260.

Wilson, M. (2002). Six views of embodied cognition. *Psychonomic Bulletin & Review*, *9*(4), 625–636.

Wolpert, D. M., Doya, K., & Kawato, M. (2003). A unifying computational framework for motor control and social interaction. *Philosophical Transactions of the Royal Society of London - Series B: Biological Sciences*, *358*(1431), 593–602.

Zhang, I., Givvin, K. B., Sipple, J. M., Son, J. Y., & Stigler, J. W. (2021). Instructed hand movements affect students' learning of an abstract concept from video. *Cognitive Science*, *45*(2), e12940.

Zhang, Z., Lei, Y., & Li, H. (2016). Approaching the distinction between intuition and insight. *Frontiers in Psychology*, *7*, 1195.

Supporting Information

Additional supporting information may be found online in the Supporting Information section at the end of the article.

Appendix A: The appendix gives the items used to assess Geometry Interest.

Appendix B: Coding criteria for insight, insight and proof

Appendix C: Categories of student responses when asked if they saw a relationship between their motions and the geometry conjectures at the end of the experiment.