

THE UNIVERSITY OF CHICAGO

GESTURE AS A BRIDGE BETWEEN NON-DECLARATIVE AND DECLARATIVE
KNOWLEDGE

A DISSERTATION SUBMITTED TO
THE FACULTY OF THE DIVISION OF THE SOCIAL SCIENCES
IN CANDIDACY FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

DEPARTMENT OF PSYCHOLOGY

BY

CASEY HALL

CHICAGO, ILLINOIS

DECEMBER 2020

TABLE OF CONTENTS

ACKNOWLEDGEMENTS	iv
LIST OF TABLES.....	vi
LIST OF FIGURES	vii
ABSTRACT.....	x
CHAPTER ONE: INTRODUCTION.....	1
Implicit and Explicit Learning in the Cognitive Literature	2
The Role of Gesture in Learning	7
Bringing the Literatures Together: Implicit Learning, Explicit Learning, and Gesture	10
Gesture as a Bridge between the Two Knowledge Types: The Current Work.....	17
CHAPTER TWO: GESTURE IN IMPLICIT AND EXPLICIT LEARNING CONTEXTS.....	26
Introduction.....	26
Method	36
Results.....	43
Discussion	61
CHAPTER THREE: DOES GESTURE DEPEND ON SPEECH TO BENEFIT LEARNING? 66	66
Introduction.....	66
Method	69
Results.....	73

Discussion	86
CHAPTER FOUR: VARYING GESTURAL KNOWLEDGE AND SPEECH 91	
Introduction.....	91
Method	96
Results.....	98
Discussion.....	112
CHAPTER FIVE: GENERAL DISCUSSION AND FUTURE DIRECTIONS 117	
REFERENCES	128
APPENDICES	139
Appendix A. Study 1 Training and Posttest Instructions.....	139
Appendix B. Study 1 Questionnaire	142
Appendix C. Study 2 and 3 Rule Video Images and Scripts	143

ACKNOWLEDGEMENTS

I have many people to thank for helping me at various stages of this dissertation, and for providing support throughout my graduate career. First and foremost, I would like to thank my advisor, Susan Goldin-Meadow, for your mentorship, guidance, and patience as this project evolved. Despite facing challenges in the early stages of my research program and in recent months, Susan helped me stay focused on my broader research questions and pushed me to explore these questions in different ways. I am so grateful for the six years I spent working with you, and for your invaluable feedback in shaping my dissertation.

I must also thank my other committee members, Susan Levine and Greg Norman, and my reader, Susan Wagner-Cook, for providing constructive feedback and academic support at every stage of my dissertation, from designing my experiments to analyzing my data. I am sincerely grateful for your intellectual curiosity and encouragement.

This work literally would not have been possible without the hundreds of participants who completed my studies online. I thank each of you for your contribution to scientific research and for making it possible for me to finish my dissertation while working remotely. I also want to thank every member of the Goldin-Meadow Lab who provided feedback during my numerous lab meetings on my dissertation studies. Your insights made my work better than I ever could have imagined. Thank you to Liz Wakefield for all you taught me about conducting excellent research during my first phase of grad school. I am honored to have collaborated with you. A special thanks to Alyssa Kersey, Michelle Hurst, and Cristina Carrazza for your help with the data analysis portion of my dissertation. Your advice gave me confidence in my work and helped make me a better researcher.

I would not have completed graduate school without the emotional support of my graduate student cohort and Mimi Halpern Maduff. Thank you all for helping me navigate both academic and personal challenges. To all my friends in the Chicago comedy scene, thank you for creating a space for me to express myself and for sharing your talents with me. To Tony Orlando and Tony Wood, thank you for all the wine-filled weekends that helped me recharge when I needed it most.

Last, but certainly not least, I want to thank my family. My parents, Tom and Jean, and my siblings, Kyle and Bobbie, for your unconditional love and constant encouragement. My grandparents, Joan and Bill and Joy and Clark, for providing every member of our family with a foundation to achieve success. Without your hard work and dedication, I would not have reached this milestone. To my cat, Rudy, for your warm snuggles after long writing sessions. And finally, to my Aunt Jeanne, our family's first PhD who laid the groundwork for me and whose passion for knowledge inspired me to reach the finish line. Thank you all.

LIST OF TABLES

Table 1. The rules for combining symbols in the system from Kaminski et al. (2008)	39
Table 2. Example of training trial.....	40
Table 3. Examples of posttest items with correct answers outlined in red.....	42
Table 4. Mean and standard deviation of posttest accuracy by problem type by learning type and gesture type; t-tests comparing accuracy on each problem type to chance performance (0.33 of problems correct; * indicate performance significantly above chance).....	54
Table 5. Followup accuracy on each problem type by learning type and gesture type – Mean (Standard Deviation).....	60
Table 6. The rules for combining symbols in the system from Kaminski et al. (2008) with updated symbols.....	71
Table 7. Examples of problem types (solutions appear in red).....	73

LIST OF FIGURES

Figure 1. Training accuracy by learning type and gesture type. Proportion correct out of 96 items. Error bars represent standard error.	46
Figure 2. Training accuracy across training blocks by learning type and gesture type. Proportion correct out of 24 items per block. Error bars represent standard error.	48
Figure 3. Training reaction time by learning type and gesture type. Error bars reflect standard error.	50
Figure 4. Training reaction time by learning type only (gesture conditions collapsed across learning type).	50
Figure 5. Posttest accuracy on all problem types combines by learning type and gesture type Note: models were run at the trial level; here, we convey accuracy at the mean level by proportion correct out of 63 items (trials). Error bars represent standard error.	53
Figure 6. Posttest proportion correct on each problem type by learning type and gesture type. Error bars reflect standard error.	54
Figure 7. Posttest proportion correct on each block (out of 21 problems) by learning type and gesture type. Error bars reflect standard error.	56
Figure 8. Posttest RT by problem type and learning type. Error bars reflect standard error.	58
Figure 9. Average reaction time per item on each posttest block by learning type and gesture type.	59
Figure 10. Posttest versus followup accuracy on each problem type by learning type and gesture for 208 participants who completed both assessments.	60
Figure 11. Histogram of proportion correct at pretest.	75
Figure 12. Pretest accuracy by problem type and condition.	76

Figure 13. Pretest reaction time by problem type and condition.....	77
Figure 14. Training accuracy by condition.....	78
Figure 15. Training performance on each training trial by condition.....	79
Figure 16. Training reaction time by condition.....	80
Figure 17. Training reaction time on each training trial by condition.....	81
Figure 18. Posttest accuracy by problem type and condition.....	82
Figure 19. Posttest reaction time by problem type and condition.....	83
Figure 20. Pretest vs. posttest accuracy by problem type and condition.....	84
Figure 21. Pretest vs. posttest reaction time by problem type and condition.....	86
Figure 22. Histogram of proportion correct at pretest.....	99
Figure 23. Pretest accuracy by problem type and condition.....	101
Figure 24. Pretest reaction time by problem type and condition.....	102
Figure 25. Training accuracy by condition.....	103
Figure 26. Training accuracy on each training trial by condition.....	104
Figure 27. Training reaction time by condition.....	105
Figure 28. Training reaction time on each training trial by condition.....	106
Figure 29. Posttest accuracy by problem type and condition.....	108
Figure 30. Posttest reaction time by problem type and condition.....	109
Figure 31. Pretest vs. posttest accuracy by problem type and condition.....	111
Figure 32. Pretest vs. posttest reaction time by problem type and condition.....	112
Figure 33. Participants in all conditions also received the following instructions prior to beginning training block 1.....	140

Figure 34. Participants in all conditions received the following instructions prior to beginning the posttest.	141
Figure 35. Video 1 – “Rule 1: The order of the symbols on the left does not change the result. For example, triangle combined with square makes triangle, and square combined with triangle also makes triangle.”	143
Figure 36. Video 2 – “Rule 2: When any symbol combines with square, the result will always be <i>that</i> symbol. For example, square combined with triangle makes triangle, and circle combined with square makes circle.”	143
Figure 37. Video 3 – “Rule 3: Circle combined with triangle makes square.”.....	144
Figure 38. Video 4 – “Rule 4: Circle combined with circle makes triangle.”	144
Figure 39. Video 5 – “Rule 5: Triangle combined with triangle makes circle.”	144

ABSTRACT

Learners can benefit from seeing an instructor gesture while explaining various concepts (e.g., Singer & Goldin-Meadow, 2005). Like non-declarative knowledge, gestures are often processed unconsciously (e.g., Goldin-Meadow, Alibali & Church, 1993). However, gesture is also seamlessly integrated with speech, a vehicle for declarative (or consciously verbalizable) knowledge (Kendon, 1980; McNeill, 1992). Gesture may partially benefit learning by representing non-declarative knowledge at the same time as speech, a representation of declarative knowledge, and thus influence a learner's developing declarative knowledge. I explore how seeing gestures may help learners transition from a state of non-declarative (or implicit) knowledge to a state of declarative (or explicit) knowledge of a novel math concept (Study 1). I also examine how the type of information conveyed uniquely in gesture interacts with information conveyed in speech to benefit learning of this same concept (Studies 2 and 3). My most intriguing findings suggest that the context in which gesture is presented matters for learning – namely, gesture promotes learning better when accompanied by explicit, declarative instruction as opposed to implicit, non-declarative instruction.

CHAPTER ONE: INTRODUCTION

People learn different types of information in different ways. The cognitive literature identifies many different learning systems, including *implicit* learning and *explicit* learning (for reviews, see Seger, 1994, and Roediger & McDermott, 1993, respectively). Each of these systems uniquely supports the acquisition of different types of knowledge and memories.

Implicit learning, learning of complex information without awareness of what has been learned, results in *non-declarative knowledge*, which is not easily articulated. Implicit learning can be distinguished from *explicit learning*, in which a learner engages in consciously-accessible, effort-based learning processes. Though implicit and explicit learning have been studied in many different contexts, questions remain as to how these processes function under different conditions and for acquiring different types of knowledge. Many researchers believe that implicit and explicit learning are not totally distinct, but represent a continuum of more or less conscious processes (Berry, 1994; Reber, 1993), and that both systems can work together depending on task demands (Knowlton, Squire, & Gluck, 1994; Mathews et al., 1989; Gluck & Bower, 1988).

The relationship between implicit and explicit learning, and the types of knowledge that result from each, is similar to the relationship between *gesture* and *speech*. Like *non-declarative* knowledge, gesture is often produced by speakers and processed by observers unconsciously (Goldin-Meadow, Alibali & Church, 1993; Garber, Alibali & Goldin-Meadow, 1998; Alibali, Flevares & Goldin-Meadow, 1997). In contrast, speech is a vehicle for *declarative knowledge*, as verbalizing knowledge requires intentional effort by a speaker, and a listener must engage in conscious processes in order to glean meaningful information from speech. Just as implicit and explicit learning processes can work together in nuanced ways to create knowledge, gesture and speech form one coherent system of communication that facilitates a learner's comprehension

(Kendon, 1980; McNeill, 1992). Co-speech gestures can benefit learning of numerous concepts when presented in instructional contexts, even when gesture and speech do not convey the same information (e.g., Singer & Goldin-Meadow, 2005; Ping & Goldin-Meadow, 2008; Valenzano, Alibali, & Klatzky, 2003).

In this introduction, I first summarize key findings from both the cognitive and gesture learning literatures. I then draw parallels between these findings, which form the foundation for my research questions and hypotheses. Finally, I describe the studies that I have designed to address these questions and summarize my main hypotheses and predictions.

Implicit and Explicit Learning in the Cognitive Literature

Implicit learning results in unconscious, *non-declarative knowledge* that cannot be verbalized. Examples of implicit learning capacities include formation of habits, preferences, and skills; priming; and simple types of memory like habituation and sensitization (Reber, 2013). *Explicit learning*, in contrast, requires conscious effort on the part of the learner. Explicit learning results in *declarative knowledge*, such as vocabulary, facts, and events. These two types of learning both stem from and contribute to the *non-declarative* and *declarative memory systems*, which depend on activation of different brain areas: declarative memory on the medial temporal lobe (MTL) and various cortical regions, particularly the prefrontal cortex and the parietal lobes (Gabrieli, 1998; Scoville & Milner, 1957; Squire, 2004; Squire & Zola, 1996); nondeclarative memory on processing areas such as the basal ganglia, cerebellum, and neocortex (e.g., Eichenbaum & Cohen, 2001; Gabrieli, 1998; McClelland, McNaughton, & O'Reilly, 1995; Reber, 2013; Squire, 2004; Squire & Zola, 1996).

Research demonstrating evidence for the existence of these two separate memory systems stems from Scoville and Milner's (1957) observations of an amnesiac patient known as H.M.,

who showed complete loss of declarative memory following surgery for epilepsy. Despite his declarative memory impairment, H.M. demonstrated learning on other memory-based tasks, such as mirror drawing and the Tower of Hanoi puzzle (Cohen, Eichenbaum, Deacedo, & Corkin, 1985; Corkin, 1968; Milner, Corkin, & Teuber, 1968). These findings were replicated in subsequent research showing that amnesiac patients can achieve normal performance on many other memory tasks, including artificial grammar learning, probabilistic category learning, and perceptual priming (e.g., Cohen & Squire, 1980; Goshen-Gottstein, Moscovitch, & Melo, 2000; Graf & Schacter, 1985; Keane et al., 1997; Knowlton & Squire, 1993, 1994, 1996; Nissen & Bullemer, 1987; Reber & Squire, 1994, 1998). Learning in these tasks demonstrates the formation of new memories while not requiring the conscious retrieval of any knowledge, and is thus implicit (Squire & Zola, 1996).

Implicit learning as a field of study began in 1967 with the work of Arthur Reber, who pioneered the modern artificial grammar learning (AGL) paradigm. Reber demonstrated that when shown a series of letter strings generated by a complex set of grammar rules and asked to memorize them, participants could later classify novel letter strings as valid or invalid above chance levels. Importantly, participants were *not* told that the strings followed a complex set of rules, indicating that they unconsciously abstracted information from their stimulus environment and later applied their non-declarative knowledge. The phenomenon of implicit learning has since been widely studied using different paradigms, including AGL (e.g., Mathews et al., 1989; Knowlton & Squire, 1994), sequence learning (e.g., Nissen & Bullemer, 1987; Lewicki, Hill, & Bizot, 1988; Reed & Johnson, 1994; Destrebecqz & Cleeremans, 2001), dynamic systems learning (e.g., Berry & Broadbent, 1984; Stanley, Mathews, Buss, & Kotler-Cope, 1989), probability learning and classification (e.g., Millward & Reber, 1968, 1972; Reber &

Millward, 1965, 1968, 1971; Knowlton et al., 1994), stereotyping and prejudice (e.g., Olson & Fazio, 2006; Plant, Peruche, & Butz, 2005), and more recently, learning of mathematical principles (Prather, 2012; Ziegler, Edelsbrunner, & Stern, 2017). These paradigms differ in many ways (e.g., in demands on perceptual and motor processing, use of auditory versus visual stimuli, and engagement of linguistic versus nonlinguistic processing), but all of them define “learning” as the participants’ ability to extract regularities in input they are exposed to without explicit instruction (Batterink, Paller, & Reber, 2019).

In addition to exploring implicit learning in different paradigms, research has also attempted to specify the mechanisms underlying implicit learning by establishing functional dissociations between implicit and explicit learning (Cleeremans, Destrebecqz, & Boyer, 1998). In comparing implicit and explicit learning, these studies manipulate factors such as type of instruction (Reber, 1976), structural salience of the stimuli (Reber et al., 1980), learning task demands (Shanks, Johnstone, & Staggs, 1997), assessment of learned knowledge (Merikle & Reingold, 1991), availability of attentional resources during learning (Cohen, Ivry, & Keele, 1990; Frensch, Buchner, & Lin, 1994; Stadler, 1995), or some combination of these factors. For example, Reber (1976) trained participants on a complex set of artificial grammar rules by presenting them with grammatically-generated letter strings and asking them to recreate the strings from memory. Participants were then tested on their knowledge of the grammar by viewing several examples of correct and incorrect strings and determining whether or not each string was valid. Importantly, at the beginning of the study, participants were given only one of two possible types of instruction: neutral, *implicit instruction* (e.g., “Look at the letter strings, memorize them, and try to recreate them”), or *explicit instruction* which directed subjects to search for the complex rules that determined the letter strings. Participants who received the

explicit instruction performed worse at both training and test than those who received implicit instruction.

Findings from similar comparative studies have found that, with the exception of priming, implicit learning mechanisms are *unconscious* (Eichenbaum & Cohen, 2001; Reber, Knowlton, & Squire, 1996), occur gradually (Eichenbaum & Cohen, 2001; Henke, 2010), and tend to be stimulus-specific (Reber et al., 1996; Eichenbaum & Cohen, 2001; Bayley & Squire, 2002). Implicit and explicit learning also differ in their facilitation of *transfer* of knowledge. Transfer, or *generalization*, is the application of learned knowledge to a task different from the original learning context. Evidence for transfer of non-declarative knowledge is somewhat weak, with many AGL studies demonstrating no ability to transfer or a transfer decrement when participants are exposed to strings constructed from one letter set and later tested on their ability to classify strings constructed from a different letter set (Brooks & Vokey, 1991; Whittlesea & Dorken, 1993; Gomez & Schvaneveldt, 1994; Seger, 1994; Dienes & Berry, 1997; see also Altmann, Dienes, & Goode, 1995, for an extension of these findings to transfer between different modalities). In contrast, explicit learning is particularly useful for forming flexible *relational representations* that can be generalized to novel contexts (Reber et al., 1996; Cohen & Eichenbaum, 1993; Eichenbaum & Cohen, 2001). Additionally, declarative knowledge acquired through explicit learning can often be verbally reported by the learner (though this is not a necessary feature of declarative memory, as not all declarative memories are easily verbalizable; e.g., Clayton & Dickinson, 1998; Fortin, Wright, & Eichenbaum, 2004). In contrast, non-declarative knowledge is *not* accessible to verbal report, and is therefore strictly unspoken (e.g., Eichenbaum & Cohen, 2001).

Though each memory system and its associated learning system function predominately for acquiring and recalling certain types of knowledge, these systems can interact in interesting ways (Knowlton, Squire, & Gluck, 1994; Mathews et al., 1989; Gluck & Bower, 1988) and individuals may vary in the system they use when approaching a task (Gluck, Shohamy, & Myers, 2002). For example, Mathews and colleagues (1989) taught participants a biconditional grammar made up of several rules (e.g., if there is an *A* in the first position, there must be an *X* in the fifth position). Unlike finite-state grammars used in previous AGL studies, the biconditional grammar was designed to generate strings that would be perceptually dissimilar to each other, so as to make implicit learning of family resemblance among exemplars particularly difficult. Participants were trained on letter strings generated by the grammar through two different types of tasks comprised of many trials: an implicit *match* task, an explicit *edit* task, or a combination condition (*match* task for the first half of trials and *edit* task for the second half of trials). For the match task, participants were told that on each trial, they would see a letter string and must remember it, and then were asked to select the string from several choices of strings. For the edit task, participants were told that on each trial, they would see a letter string generated by a complex set of rules and each string would have one to four letters that were incorrect, and that they must try to explicitly discover the rules of the grammar by marking the letters they thought were incorrect on each string. Participants in all training conditions received the same posttest, in which they were shown several strings and asked to select the grammatically-correct string over many trials (half the strings were previously seen during training, half were new). At posttest, the combination condition performed best, followed by the explicit condition, with the implicit condition performing worst. In discussing these findings, the authors proposed that finite-state grammars used in previous AGL studies convey a family-resemblance structure that can be

learned implicitly; in contrast, the biconditional grammar generated strings that were perceptually dissimilar to each other, so as to make learning of family resemblance among exemplars through implicit processes alone particularly difficult. Based on the success of the implicit followed by explicit group, the authors further argued that learning of some rule-based tasks may benefit from a synergy between implicit and explicit learning modes, in which participants first develop an implicit knowledge base *before* generating an explicit model of the concept to be learned. Even though participants in the implicit→explicit condition may have been unable to fully extrapolate the rules underlying the biconditional grammar through the match task, the familiarity with grammatical strings gained through the match task may have provided a strong foundation on which to build declarative knowledge of the rule structure.

The Role of Gesture in Learning

Implicit and explicit learning mechanisms and their associated knowledge types have been explored in different contexts. However, few studies have directly addressed the parallels between these two learning systems and the gesture-speech communication system. Gesture's influence on learners' behavior is often not consciously accessible or verbalized (Goldin-Meadow, Alibali & Church, 1993; Garber, Alibali & Goldin-Meadow, 1998) and is thus implicit. At the same time, gesture is seamlessly integrated with speech, a vehicle for declarative knowledge (Church, Kelly & Holcombe, 2014), and thus, gesture may interact with explicit learning processes to benefit learning.

Gesture can facilitate a student's learning of a concept, whether the student performs their own gestures (Cook, Mitchell, & Goldin-Meadow, 2008; Goldin-Meadow, Cook, & Mitchell, 2009) or observes a teacher's gestures (Singer & Goldin-Meadow, 2005). This dissertation will focus on learning through *seeing* gesture in instructional contexts. Experimental studies in which

gesture instruction is compared to instruction with speech alone have shown that observing co-speech gesture can play a causal role in improving learners' understanding of numerous concepts, including mathematical equivalence (e.g., Church, Ayman-Nolley, & Mahootian, 2004; Singer & Goldin-Meadow, 2005; Cook & Goldin-Meadow, 2006), Piagetian conservation (Ping & Goldin-Meadow, 2008), symmetry (Valenzano, Alibali, & Klatzky, 2003), and word learning (Capone & McGregor, 2005; Goodrich & Hudson Kam, 2009; Mumford & Kita, 2014, Wakefield, Hall, James, & Goldin-Meadow, 2018). Furthermore, the positive effects of learning through seeing gesture persist over time (Cook, Duffy, & Fenn, 2013; Wakefield et al., 2019; Congdon et al., 2017).

It is likely that many different mechanisms contribute to gesture's positive effects on learning, including its ability to draw attention to relevant stimuli in a learning environment (Shimpi & Huttenlocher, 2007; Rader & Zukow-Goldring, 2010). Observed gestures may also help learners follow along with a teacher's speech (Wakefield et al., 2018b). Gesture itself is largely unspoken (though see Shintel, Nusbaum, & Okrent, 2006, for an exploration of unconscious vocal modulation described as "verbal gesture"), and yet, learners can glean information from gesture when gesture is paired with meaningful speech. For example, Wakefield and colleagues (2018) used eyetracking technology to measure how children allocated their visual attention when watching a video of a teacher explaining how to solve a math equivalence problem (e.g., $5 + 6 + 3 = \underline{\hspace{1cm}} + 3$). In the speech + gesture video condition, the teacher said, "I want to make one side equal to the other side" (an *equalizer* strategy) while simultaneously performing a *grouping* gesture (point to the 5 and 6 on the left side of the equation using a v-handshape, followed by a point to the blank on the right side). In the speech alone condition, the speaker only said the equalizer strategy without gesturing. Children in the

speech + gesture condition performed better than children in the speech alone condition on a posttest of novel math equivalence problems. The authors also found that children in the speech + gesture condition looked more to the problem and looked less to the instructor when watching the video, and were better at pairing their visual attention with the instructor's speech. These looking patterns positively predicted learning at posttest, but they did *not mediate* the effects of training condition on posttest performance. However, training condition *moderated* the effect that *following along with speech* had on learning—*following along with speech* predicted learning for children in the gesture condition, but not for children in the speech alone condition. Thus, these results suggest that gesture impacts how learners attend to a teacher's speech and in turn, what learners gain from that speech.

Gesture can also convey different kinds of information; *deictic* (pointing) gestures can refer to objects in the environment, while *iconic* gestures can convey physical attributes (e.g., placing a hand high above the head to convey the height of a tall person) or actions (e.g., quickly moving an empty fist forward to depict throwing). The *type of information* conveyed in gesture in a learning context is crucial for gesture to have a positive impact on learning. For example, Goodrich and Hudson Kam (2009) used iconic gestures to teach 3- and 4-year-old children novel intransitive verbs. Children saw an event, such as a toy spinning on a turntable, and then saw an experimenter say a novel label for the event in speech (e.g., *dack*). While speaking, the experimenter performed either a relevant iconic gesture (e.g., a flat palm tracing a circular motion to illustrate the manner and path of the toy moving on the turntable) or an interactive gesture that did not disambiguate the meaning of the verb (e.g., pointing at the child); in a control condition, the experimenter labeled the event without gesturing. After the learning phase, children observed the same event they observed during learning, along with a new toy

performing a new action, and were asked, “Which toy is dacking?” Only children who had seen the *iconic* gestures during learning selected the correct action above chance, indicating that the representational information about the events conveyed in gesture helped children to correctly paired the novel verb with its action referent.

Though gesture likely benefits learning through many different mechanisms, one mechanism that has not been well explored is how gesture interacts with implicit and explicit learning mechanisms to help learners transition from a state of non-declarative to declarative knowledge. In the next section, I discuss the similarities that learning through gesture shares with each type of learning and the knowledge structures that result from each.

Bringing the Literatures Together: Implicit Learning, Explicit Learning, and Gesture

Implicit Learning and Gesture. Based on early cognitive research with amnesiac patients, recent work has further explored how gesture impacts learning in patients with memory impairment. Klooster and colleagues explored how patients with hippocampal amnesia, who have impaired declarative memory (Klooster et al., 2014), and patients with Parkinson’s disease (Klooster et al., 2015), who have impaired non-declarative memory, perceived gesture instruction when learning the Tower of Hanoi (TOH) task (these studies also examined the gestures that patients themselves produced; for the purposes of this dissertation, I focus on the findings related to learning through *observed* gesture). In a procedure adapted from Cook and Tanenhaus (2009), both healthy and memory-impaired participants observed a speech + gesture explanation of the TOH, and then solved the problem on a computer as their mouse movements were recorded. In their original study, Cook and Tanenhaus (2009) found that healthy participants’ mouse movements were affected by the gestures shown in the explanations (e.g., participants who saw an explanation with high arching gestures moved the mouse with significantly higher curvature

than participants who saw an explanation with flatter gestures). Hippocampal amnesiac patients, who have impaired *declarative* memory, performed like healthy participants (Klooster et al., 2014). In contrast, patients with Parkinson's disease, who have impaired *non-declarative* memory, moved the mouse in the same manner regardless of the arch height of the gestures in the explanations they watched (Klooster et al., 2015). Together, these results suggest that declarative memory is *not* required to benefit from the information observed in gesture, and that seeing gesture in an instructional context can support creation of new non-declarative knowledge. Next, I will elaborate on more similarities between learning through gesture and implicit learning.

Information learned through gesture can be unconscious. In addition to this evidence that gestural knowledge may be processed similarly to other types of information learned implicitly, some research suggests that gesture's influence on learners' behavior is often not consciously accessible or verbalized (Goldin-Meadow, Alibali & Church, 1993; Garber, Alibali & Goldin-Meadow, 1998). In communication, listeners rarely attend to a speaker's gestures overtly (e.g., Gullberg & Holmqvist, 2006; Gullberg & Kita, 2009) and are typically not aware of the fact that the information they gleaned from a speaker came from gesture (Alibali, Flevares & Goldin-Meadow, 1997). Like implicit learning processes, gesture perception occurs automatically and its influence on speech processing is obligatory (Kelly, Özyürek, & Maris, 2010). For example, Kelly, Özyürek, and Maris (2010) found that adult participants were faster to map action primes (e.g., a person holding a knife and chopping vegetables) to speech + gesture targets (or speech + gesture targets to action primes) when the speech + gesture pairings contained congruent information (e.g., speech: "chop"; gesture: grip handshape and chopping motion) as opposed to incongruent information (e.g., speech: "chop"; gesture: twisting motion

OR speech: “twist”; grip handshape and chopping motion). Furthermore, observing gesture can change a learner’s knowledge even if the learner is unable to verbally report that this knowledge came from gesture (Broaders & Goldin-Meadow, 2010; Gurney, Pine & Wiseman, 2013; Kelly et al., 1999; Thompson, 1995; Kelly et al., 2004). For example, Ellis Weismar and Hesketh (1993) taught children novel words through observed gestures or emphatic stress and found that children could not identify words they had learned through gesture, even though seeing the gestures had increased their subsequent comprehension of the novel words.

Gesture is rooted in motor experience. Gesture is also an action, and people gesture when they talk without conscious effort; in fact, children learning to speak often produce gestures before their first words (Iverson & Goldin-Meadow, 1998). Thus, the benefit of learning through gesture may partially arise from the lifelong experience of producing and seeing gesture. This salient feature of gesture is reminiscent of the connection between procedural non-declarative knowledge and the motor system (for reviews, see Seger, 1994; Shanks & St. John, 1994; Shanks, 2005). Ping, Goldin-Meadow, and Beilock (2014) explored how the motor system may be involved in learning through seeing gesture. Participants saw video clips of a woman speaking a sentence (e.g., “The woman hammered the nail into the wood”) while producing a gesture that conveyed additional information about the object (e.g., a gesture demonstrating hammering a vertically oriented nail or a gesture demonstrating hammering a horizontally oriented nail). Participants were then shown a picture of either a vertically oriented nail or a horizontally oriented nail and were asked whether the object in the picture had been named in the sentence. Participants were faster to respond “yes” when the picture had been preceded by a spoken sentence with a congruent gesture than with an incongruent gesture. In a second experiment, a new group of participants were asked to complete the same task while performing

movements with either their arms and hands (thus using motor resources involved in producing gesture) or their legs and feet. Only participants who performed movements with their arms and hands demonstrated interference on the sentence-processing task—that is, they did *not* respond more quickly to pictures that were congruent with the speaker’s gestures than to pictures that were incongruent. These results suggest that seeing and processing another’s gestures may rely on the listener’s own motor system, particularly those areas responsible for producing one’s own gestures.

Learning with gesture persists over time. Just as implicit learning occurs gradually over time, requiring repeated practice with a task (Eichenbaum & Cohen, 2001; Henke, 2010), learning through gesture can also emerge over time (Cook, Duffy, and Fenn; 2013; Wakefield et al., 2019; Congdon et al., 2017). For example, Cook, Duffy, and Fenn (2013) trained children on math equivalence problems by having them observe a teacher solve problems through either speech alone or through speech and gesture. They then measured children’s learning with an immediate posttest and a followup test 24-hours later that also included transfer problems. The researchers found that children who learned through speech and gesture performed better than children in the speech alone condition on both the initial posttest and the followup posttest and transfer tests 24-hours later. Importantly, the effect of condition on followup test performance was reliable even when initial posttest performance was controlled for. Thus, gesture facilitated retention of knowledge beyond initial differences in learning. Congdon and colleagues (2017) similarly found that the benefits of speech and gesture instruction on math equivalence learning emerged at a 24-hour posttest, and persisted at a 4-week posttest. The positive effects of gesture on learning over time have also been observed for word learning (Wakefield et al., 2018) and mental rotation (Wakefield et al., 2019).

Gesture can convey unique unspoken knowledge. Perhaps the strongest evidence that gesture may function through processes similar to implicit learning is the fact that gesture can provide unspoken knowledge to a listener—that is, information not found in speech. The information conveyed in a speaker’s gestures can either reiterate (match) or differ from (mismatch) the information conveyed in speech. Recall how Wakefield and colleagues (2017) presented a video lesson to children learning to solve math equivalence problems (e.g., $5 + 6 + 3 = \underline{\quad} + 3$) in which a teacher simultaneously performed different speech and gesture strategies. In an earlier study by Singer and Goldin-Meadow (2005), children were similarly taught math equivalence by observing as an experimenter performed different speech and gesture strategies. In all learning conditions, children saw and heard the experimenter give an equalizer strategy in speech: “I want to make one side equal to the other side.” In one condition, children saw and heard this equalizer strategy in speech without gesture (*Speech Alone*). In a second condition, children saw and heard the experimenter perform a gesture that conveyed the same equalizer strategy as her spoken strategy (e.g., a sweeping gesture with the left hand first under the left side of the equation and then under the right with the right hand; *Speech + Matching Gesture*). Children in a third condition saw and heard the experimenter perform gestures that conveyed a different strategy from the equalizer strategy conveyed in speech (e.g., points to the three numbers on the left side of the equation followed by a “take away” gesture under the number on the right side, an “add-subtract” strategy; *Speech + Mismatching Gesture*). In a final control group, children saw and heard the experimenter perform both the equalizer and the add-subtract strategies in speech (*Speech Alone – Two Strategies*).

Children in the Speech + Mismatching Gesture condition improved significantly more from pretest to posttest than children in all other conditions, and children in the Speech Alone –

Two Strategies condition performed significantly worse than children in all other conditions. Thus, children in the Speech + Mismatching Gesture condition did not simply perform better because they received two strategies instead of one, but rather because one of these strategies was conveyed uniquely in gesture. This finding suggests that unspoken knowledge conveyed by a teacher's gesture may make the teacher's spoken knowledge more salient to a learner, even when gesture conveys information that is different from speech. In this way, gesture may serve as a vehicle for conveying non-declarative knowledge.

Explicit Learning and Gesture. Though learning through gesture shares many characteristics of implicit learning, gesture is also linked with explicit learning processes.

Gesture is seamlessly integrated with speech, a vehicle for declarative knowledge.

Unlike other types of movements, gesture is seamlessly integrated and routinely occurs with speech, which is the gold standard for declarative knowledge (Bernardis & Gentilucci, 2006; Church, Kelly, Holcombe, 2014; Loehr, 2007; McNeill, 1992). It has long been proposed that gesture and speech form one coherent system of communication, as these modalities are synchronized temporally (Nobe, 2000) and in the information they convey (McNeill, 1992).

Gesture's relationship with speech can facilitate comprehension of that speech (McNeil, Alibali & Evans, 2000; Kelly, McDevitt, & Esch, 2009), and vice versa (Kelly, Özyürek, & Maris, 2010). A study by Eliza Congdon and colleagues (2017) directly explored how the *synchrony* between speech and gesture impacts children's learning of math equivalence. Children were instructed on problems in one of three learning conditions in which an experimenter varied the timing of speech and gesture: the first was a control condition in which the experimenter produced two strategies in speech, an equalizer strategy and an add-subtract strategy (speech followed by speech or S→S); in the second condition, an equalizer strategy in speech preceded

an add-subtract strategy in gesture (speech followed by gesture or S→G); in the third condition, the equalizer strategy in speech was performed at the same time as the add-subtract strategy in gesture (simultaneous speech and gesture or S+G). After being trained on problems in one of these conditions, children took an immediate posttest, a followup posttest 24-hours later, and an additional followup 4 weeks later. There were no significant differences on performance at the immediate posttest, but a significant interaction emerged between time point and condition on both of the followup posttests: the simultaneous speech and gesture condition performed better than the other two conditions. The simultaneous condition also improved from immediate to next-day posttest, and was retained at the 4-week followup, whereas performance in the other two conditions did not improve, and actually worsened at some time points.

These results suggest that gesture's simultaneous presentation with speech is important for gesture to have an impact on learning. As was observed by Singer and Goldin-Meadow (2005), presenting children with two different strategies in this study was not sufficient to benefit learning, as there was no difference in performance between the S→S and S→G conditions. Crucially, even presenting two strategies in different modalities was insufficient if these strategies were not presented simultaneously. Gesture's benefit on learning when it is presented simultaneously with speech provides evidence that part of the mechanism through which gesture's relationship with speech impacts learning is through its ability to seamlessly provide a learner with unspoken (non-declarative) knowledge at the same time as speech (declarative knowledge).

Learning with gesture promotes generalization of learned knowledge. In addition to forming an integrated system with speech, gesture shares another important quality with declarative knowledge. Just as knowledge acquired through explicit learning processes can be

flexibly applied to novel problems (Reber et al., 1996), learning through seeing gesture can also facilitate generalization ability (Cook, Duffy, & Fenn, 2013; Congdon et al., 2017; Wakefield et al., 2018; though see Beilock & Goldin-Meadow, 2010 for evidence that in some tasks, gesture may tie learners to specific characteristics of the learning environment). Cook, Duffy and Fenn (2013) showed children videos of an experimenter teaching children to solve math equivalence problems through either an equalizer strategy in speech alone or an equalizer strategy in speech with a matching gesture. Children in the gesture condition showed increased success on generalization problems on a posttest given 24-hours later, compared to children in the speech alone condition, even when controlling for differences in initial learning. In this study, the information in gesture matched the equalizer principle explained in speech, which may have been particularly helpful for helping children learn to generalize their knowledge to novel problem structures. However, similar studies have demonstrated benefits on generalization after seeing mismatching gestures that convey a problem-solving algorithm, such as grouping or add-subtract (e.g., Wakefield et al., 2017; Congdon et al. 2017). Together, these findings suggest that gesture may benefit generalization regardless of the *type* of information conveyed in gesture, as long as the non-declarative knowledge complements speech in a meaningful way.

Gesture as a Bridge between the Two Knowledge Types: The Current Work

In this dissertation, I explore implicit and explicit learning through the lens of gesture by focusing on the similarities that learning through gesture shares with both learning processes. I propose that gesture may interact with implicit and explicit learning processes to help learners transition from a state of non-declarative to declarative knowledge of a concept. I argue that, as part of this process, gesture benefits learning by representing non-declarative knowledge at the same time as speech, a representation of declarative knowledge, and thus influences a learner's

developing declarative knowledge. In exploring this hypothesis, I also aim to both extend the implicit and explicit learning literature to a new domain (math learning) and to clarify the mechanisms through which gesture benefits learning across the lifespan (by extending previous work with gesture and math learning in children to adult learners).

Study 1, Question 1: How do implicit and explicit learning processes function separately or together during instruction of a symbolic math system? In Study 1, I manipulate adult participants' *intention to learn* the rules of a symbolic math equivalence system from Kaminski, Sloutsky, and Heckler (2008). This study extends AGL paradigms used in the learning literature to a novel task that can easily incorporate gesture. The symbolic math equivalence system I use is similar to both finite-state and biconditional artificial grammars used in previous work (e.g., Reber, 1976; Mathews et al., 1989). Like a finite-state grammar, the symbol system can create multiple similar exemplars (in this case, equations). At the same time, these equations are constructed according to conditional rules (e.g., two symbols combine to form exactly one other symbol). I train participants on equations in different learning conditions, in which I vary whether participants are told to simply remember the equations (implicit learning), to try to decipher the rules of the symbol system (explicit learning), or to follow implicit instructions for the first half of learning and to follow explicit instructions for the second half (implicit→explicit learning). I then test participants on both their ability to recognize equations seen during training, as well as their understanding of the underlying rule structure of the equations, and I examine whether learning persists on a 24-hour followup test. This experiment thus tests how intention to learn (unconscious and passive vs. conscious and effortful), which distinguishes implicit and explicit learning, functions in a math-learning paradigm. Additionally, by including an implicit→explicit condition, I explore how well

participants learn the symbol system when they first establish an implicit knowledge base before engaging in explicit learning processes.

My learning conditions in Study 1 are closely modeled after AGL paradigms comparing implicit and explicit learning (e.g., Reber, 1976; Mathews et al., 1989). Early studies comparing the effect of implicit vs. explicit instruction on AGL found that participants given implicit instruction perform better at posttest than participants given explicit instruction (Reber, 1976; Brooks, 1978). However, subsequent research has found that key changes to the experimental paradigm can reverse these findings. For example, when grammatical exemplars are presented in a structured, rather than random display, participants given explicit instructions to search for grammatical rules outperform participants given implicit instructions to simply remember the exemplars (Reber et al., 1980). Similarly, Mathews and colleagues (1989) found that explicit instruction can be more beneficial than implicit instruction when learning a complex biconditional grammar. Berry and Broadbent (1988) extended these findings to a dynamic systems task, in which manipulating the “structural salience” of the task impacted participants’ ability to learn from implicit compared to explicit instruction. My paradigm uses graphic shapes to makeup equations rather than traditional letter strings and the task is framed in terms of “solving” equations. These cues paired with explicit instruction to search for rules may be particularly effective for learning (though previous work has shown that participants can use graphic symbols as classification cues for associated grammatical letter strings even when given implicit instructions; see Morgan, Meier, & Newport, 1987 and Altmann et al., 1995). My first research question for Study 1 thus concerns whether implicit and explicit learning processes, here defined as “unintentional vs. intentional learning,” can be applied to a novel task not traditionally used in the implicit vs. explicit learning literature.

Study 1, Question 2: Does gesture interact with implicit and explicit learning

processes to benefit learning? In this same experiment, I will also explore how unspoken, non-declarative knowledge conveyed in gesture interacts with intention to learn by presenting the implicit vs. explicit training conditions with or without gesture. This study design results in six different learning conditions: implicit alone, implicit + gesture, explicit alone, explicit + gesture, implicit→explicit alone, and implicit→explicit + gesture. To keep the six conditions comparable, participants in all conditions were shown identical videos of a teacher looking at an equation and saying a neutral phrase in speech, “Take a look at this problem.” In the gesture conditions, the teacher simultaneously produced a grouping gesture below the equation while speaking. This gesture is adapted from previous studies on the role of gesture in learning of math equivalence in children and depicted a problem-solving strategy that also highlighted the two-symbol combinations that make up the rules of the system.

My second research question for Study 2 thus asks how gesture will interact with implicit and explicit learning paradigms. One possibility is that by conveying non-declarative knowledge, gesture will simply benefit learning regardless of the context in which it is presented. Non-declarative knowledge can influence learning by guiding early processing of stimuli, rapidly providing new data to be processed by the declarative system (Musen & Treisman, 1990). By providing a visual representation of non-declarative knowledge, seeing gesture may make learners more likely to attend to task-relevant information. In attending to this information, non-declarative representations may provide the declarative system with the information necessary to construct a correct declarative representation. If this is the case, I would expect to see an overall effect of gesture instruction on learning across conditions.

Another possibility is that gesture must be paired with meaningful co-speech in order to benefit learning. In other words, learners may struggle to glean non-declarative knowledge conveyed in gesture if this knowledge is not grounded in declarative knowledge conveyed in speech. If this is true, gesture may not facilitate learning in any condition.

A final possibility is that gesture may differentially impact learning depending on the context it is presented in. For example, if gesture benefits learning by serving as a vehicle for non-declarative knowledge that also seamlessly integrates with declarative knowledge, gesture may better facilitate learning when it is presented in an explicit learning context than when it is presented in an implicit learning context. In this case, gesture would be grounded in an explicit context, allowing learners to benefit from the unspoken knowledge conveyed in gesture while consciously trying to learn how to solve the problems. In contrast, when gesture is presented in an implicit learning context, learners may not fully engage with the cognitive processes necessary to take advantage of the information conveyed in gesture. In this case, I would expect to see an interaction of gesture instruction (gesture present vs. gesture absent) with learning type (implicit vs. explicit). Lastly, if gesture is truly functioning as a bridge between implicit and explicit learning processes, it may be most effective in the implicit → explicit condition. Including the three gesture conditions allowed me to test whether gesture's effectiveness as a learning tool depends on the instructional context in which it is presented—specifically, whether learners view gesture while engaging in either implicit or explicit learning processes, or a combination of both.

Study 2: Are observed gestures' impact on learning dependent on gesture's relationship with speech? In Study 2, I ask whether the unspoken, non-declarative knowledge conveyed in gesture is optimally useful for learning when it is paired with declarative

knowledge—that is, meaningful speech. If gesture benefits learning by depicting non-declarative knowledge that is also seamlessly integrated with speech, a vehicle for declarative knowledge, then gesture may be less beneficial for learning when it is presented without speech, or with vacuous speech. To test this hypothesis, I train participants to solve novel symbolic math equivalence problems in the same system used in Study 1 in one of three learning conditions: *declarative knowledge alone* (learners see a teacher explain problems using an equalizer strategy in speech alone), *non-declarative knowledge alone* (learners see a teacher produce an equalizer gesture with neutral speech: “Take a look at this problem”), or *simultaneous declarative and non-declarative knowledge* (learners see a teacher produce an equalizer strategy in speech with a matching gesture). Thus, I will test 1) whether gesture is more effective for learning when it reiterates spoken, declarative knowledge, as opposed to just providing unspoken, declarative knowledge, and 2) whether the non-declarative knowledge provided by instruction with gesture alone is more, less, or equally effective for learning than speech alone instruction.

Though part of gesture’s ability to promote learning may stem from its unique pairing with speech, there is some evidence to suggest that instruction with gesture alone may facilitate learning in contexts in which learners have existing declarative knowledge prior to instruction. Research from math equivalence learning in children suggests that it is possible to benefit from gesture instruction *alone* (without accompanying speech) when learners are asked to *do* the gestures themselves. Cook, Mitchell, & Goldin-Meadow (2008) found that children learned to solve math equivalence problems equally well when they were instructed to either perform simultaneous speech and gesture, or to just perform a gesture, as measured by performance on a delayed posttest following training. Both gesture conditions performed better than a speech alone condition in which children were taught to say a strategy in speech, but not to gesture.

Though the children in this study did not understand how to solve math equivalence problems prior to participating in the study, they did have experience with solving other types of math problems. Unlike in Study 1, adults in Study 2 are told the rules of the symbolic math equivalence system prior to watching instructional videos with or without gesture, and thus will also have a declarative knowledge base on which they can build. Work in social psychology has shown that individuals who have discrepancies between their implicit and explicit attitudes or beliefs are particularly motivated to carefully consider new relevant information (Brinol, Petty & Wheeler, 2006; Rydell, McConnell, & Mackie, 2008). Based on this account, seeing gestures that convey information different from information conveyed in speech could create a type of cognitive dissonance, perhaps unconscious, that encourages learners to engage in increased processing of information relevant to the task. In other words, the non-declarative knowledge embodied in gesture may not itself change thinking but may, in conjunction with already held declarative knowledge, encourage learners to seek new data that will subsequently change thinking. Approaching gesture instruction with intact declarative knowledge may thus help participants glean non-declarative knowledge from the gestures they see (whether or not these gestures are paired with meaningful speech), and subsequently update their declarative knowledge of math equivalence. If this is the case, gesture alone instruction may promote learning as well as speech + gesture instruction, or at least better than instruction with speech alone.

Study 3: How does the *type* of unspoken knowledge conveyed in gesture impact learning? In Study 3, I further explore how the unspoken, non-declarative knowledge conveyed in gesture interacts with the declarative knowledge conveyed in speech to impact learning by manipulating the type of information conveyed in each modality. Specifically, I present learners

with either one or two strategies for solving novel symbolic math equivalence problems and vary the modality (speech vs. gesture) of each strategy. Most importantly, I will also compare how the *type* of information (principle vs. algorithm) conveyed by each strategy impacts learning, as well as whether the modality (speech vs. gesture) in which each strategy is presented affects learning, allowing me to further specify the context in which the unspoken knowledge provided by gesture is most beneficial for learning. This question has not been directly explored in math equivalence and gesture learning paradigms with children, but has important implications for gesture's use as a learning device in educational settings. I used adapted versions of the equalizer (principle) and grouping (algorithm) strategies from previous work in math equivalence learning in children to create 4 learning conditions: 2 matching conditions (equalizer speech + equalizer gesture; grouping speech + grouping gesture) and 2 mismatching conditions (equalizer speech + grouping gesture; grouping speech + equalizer gesture).

I expect gesture to facilitate learning regardless of whether it reiterates spoken knowledge or conveys new unspoken knowledge. Based on the finding by Singer and Goldin-Meadow (2005) that children learned to solve math equivalence problems more effectively through two strategies when one of the strategies conveyed an algorithm (add-subtract) uniquely in gesture—that is, when gesture conveyed new unspoken knowledge—I expect the equalizer speech + grouping condition to outperform both matching conditions. If gesture simply benefits learning when it conveys non-declarative that is different from speech regardless of the type of knowledge it conveys, the grouping speech + equalizer gesture condition should perform as well as the other mismatching condition. However, an interaction between the *type* of knowledge conveyed and whether gesture reiterates this knowledge or conveys new knowledge will indicate that gesture may be most useful as a learning tool when it complements the declarative

knowledge of a teacher’s speech in a specific way. Based on findings that gesture instruction can promote generalization of a learned concept (Cook, Duffy, & Fenn, 2013; Congdon et al., 2017; Wakefield et al., 2018), gesture may be particularly effective at conveying a complex, but “general” principle that can be applied across all problem types. At the same time, gesture is non-declarative, and therefore already outside of a learner’s consciousness, and thus, a gesture that strengthens the “structural salience” (Reber et al., 1980) of the task by depicting a simple algorithm may also benefit learning. The results of Study 3 distinguish between these possibilities.

In sum, my dissertation advances existing research in three key aspects: 1) In Part 1, I will explore whether “intention to learn” effectively differentiates between implicit and explicit learning processes when these processes are applied to a novel math-learning task. I will also ask whether gesture interacts with these different learning contexts, perhaps by helping learners transition from a state of implicit knowledge to one of explicit knowledge. This study has the potential to lead to better understanding of both the mechanisms supporting implicit and explicit learning, and the complex mechanisms by which gesture may support learning. 2) In Part 2, I will exploit gesture’s ability to convey unspoken knowledge in a learning context to investigate how the “unspoken” quality of non-declarative knowledge contributes to learning depending on the type of information conveyed in unspoken knowledge. 3) Part 2 will also build upon a large body of work on gesture’s positive effects on *children* learning math equivalence by exploring how gesture helps *adult* participants learn a symbolic math equivalence task. Thus, my dissertation allows for a developmental discussion of how gesture benefits learning across the lifespan.

CHAPTER TWO: GESTURE IN IMPLICIT AND EXPLICIT LEARNING CONTEXTS

Introduction

Decades of experiments on implicit and explicit learning have demonstrated that learning of complex tasks can occur through both implicit and explicit instruction (e.g., artificial grammar learning: Reber, Kassin, Lewis, & Cantor, 1980; Mathews et al., 1989; sequence learning: Nissen & Bullemer, 1987; Lewicki, Hill, & Bizot, 1988; Reed & Johnson, 1994; Destrebecqz & Cleeremans, 2001; dynamic systems learning: Berry & Broadbent, 1984; Berry & Broadbent, 1988; Stanley, Mathews, Buss, & Kotler-Cope, 1989; serial reaction time: Willingham, Nissen, & Bullemer, 1989; probability learning and classification: Millward & Reber, 1968, 1972; Reber & Millward, 1965, 1968, 1971; Knowlton, Squire, & Gluck, 1994; stereotyping and prejudice: Olson & Fazio, 2006; Plant, Peruche, & Butz, 2005; and more recently, with learning of mathematical principles: Prather, 2012; Ziegler, Edelsbrunner, & Stern, 2018). *Implicit learning*, learning without conscious awareness, results in *non-declarative knowledge*, whereas explicit learning, learning through consciously-accessible, effort-based processes results in *declarative knowledge*. Though early research on these two learning and knowledge systems often highlights their differences, many researchers believe that both systems work together in nuanced ways depending on different factors, including the type of information to be learned, task demands, and the manner in which knowledge is retrieved (Knowlton et al., 1994; Mathews et al., 1989; Gluck & Bower, 1988).

Reber (1967) first demonstrated implicit learning of an artificial grammar by asking participants to memorize a series of letter strings generated by a complex set of rules. After memorizing several strings, participants classified novel grammatical strings as valid or invalid above chance levels. Because participants were *not* told that the strings followed a set of rules,

Reber concluded that participants had acquired the grammatical rules implicitly, and that this unconscious knowledge is encoded in *nondeclarative* memory as an abstract representational system (Reber, 1967). Though Reber's original study examined implicit learning in isolation, subsequent studies adapted Reber's artificial grammar learning (AGL) paradigm and replicated the original result when comparing the effect of implicit learning instructions to explicit learning instructions, in which participants were explicitly told to search for rules. Participants who received implicit instructions to simply memorize strings performed better on a subsequent string-classification task compared to participants who received explicit instructions to search for grammatical rules (Reber, 1976; Reber & Lewis, 1977; Brooks, 1978). Other work exploring the implicit and explicit learning distinction found that explicit instruction could be as beneficial, or in some cases, *more* beneficial than implicit instruction depending on the presentational structure of grammatical strings or the type of grammar to be learned. For example, Reber and colleagues (1980) found that explicit instructions to search for grammatical rules resulted in better learning than implicit instructions to simply memorize strings when the “salience” of the grammatical patterns was increased – specifically, when patterns of the letter ordering that made up the grammar were presented in a structured, rather than random, order. In other cases, even when strings were presented in a random order, subsequent classification performance was equally good after memorization instructions or rule-search instructions (e.g., Dienes, Broadbent, & Berry, 1991; Dulaney, Carlson, & Dewey, 1984; Mathews et al., 1989, Experiments 1 and 2).

Mathews and colleagues (1989) further compared the effects of implicit (memorization) and explicit (rule-search) learning processes by manipulating not only the instructions given to participants, but the learning tasks themselves. This study also explored how the implicit and explicit learning systems may work together to benefit learning. In one experiment, participants

learned an artificial grammar adapted from Reber et al. (1980) through either an implicit (match) task, an explicit (edit) task, or a combination of both tasks (either match followed by edit, or edit followed by match). For each match task trial, participants were shown a string and told to remember it, and then were asked to select the string from several choices of strings. For the edit task, participants were told at the beginning of the task that the strings they would see were generated by a complex set of rules; on each trial, they were shown an invalid string and asked to mark the letters they thought were incorrect. Classification performance of both new strings and strings seen during the learning tasks were equally good after the match task or edit task alone or in combination (regardless of order). Following this experiment, the authors replicated the same procedure with a “biconditional” grammar. In this grammatical structure, the first four letters of any string were chosen randomly as T, P, or V and the next four letters were chosen such that T predicted an X in a corresponding position, P predicted a C, and V predicted an S. Thus, for example, TPPVXCCS was a valid string. For this biconditional grammar, classification performance was best for participants who first received the implicit (match) task and then the explicit (edit) task. Classification performance for the explicit (edit) task alone and the explicit (edit) task followed by implicit (match) task were equally good, while participants who received the implicit (match) task alone performed next to chance on classification and worse than all other groups. The authors hypothesized that only the finite-state grammar involved a family-resemblance structure that could be learned implicitly, whereas the biconditional grammar was designed to generate strings that would be perceptually dissimilar to each other, so as to make implicit learning of family resemblance among exemplars particularly difficult. Based on the success of the implicit followed by explicit groups on the biconditional grammar, the authors further argued that learning of some tasks may benefit from a synergy between implicit and

explicit learning modes, in which participants first develop an implicit knowledge base *before* generating an explicit model of the concept to be learned.

The possible synergy of implicit and explicit learning processes explored by Mathews and colleagues (1989) has been investigated in other studies as well. These studies examined how learners may transition from a state of implicit knowledge to a state of explicit knowledge, as well as how implicit and explicit processes may work simultaneously to benefit learning. For example, Stanley and colleagues (1989) found that when participants completed a dynamic control task, their performance improved with repeated procedural practice (implicit learning), but participants were unable to verbalize their acquired knowledge until near the end of their training session on the task (“ability to verbalize” is a hallmark of declarative knowledge acquired through explicit learning). The authors concluded that in this task, participants first acquired nondeclarative knowledge through implicit learning prior to acquiring declarative knowledge. Similar results have been shown in artificial grammar learning (Reber & Lewis, 1977), pattern completion tasks (Bowers, Regehr, Balthazard, & Parker, 1990), and reactive sequential decision making (Sun, Merrill, & Peterson, 1998, 2001). Based on this finding that declarative knowledge can lag behind but eventually emerge along with nondeclarative knowledge, researchers have suggested that declarative knowledge can sometimes be “extracted” from nondeclarative knowledge—that is, initially implicitly acquired knowledge can continuously develop toward more explicitly available knowledge (e.g., Stanley et al., 1989; Seger, 1994; Siegler & Stern, 1998; Sun et al., 2001; Sun, Slusarz, & Terry, 2005).

More recent work has explored implicit and explicit learning processes and the knowledge bases that result from each in the context of mathematics learning. For example, Prather (2012) found that adult participants improved their knowledge of arithmetic principles

when they were asked to view and evaluate many principle-consistent and -inconsistent equations, similar to the memorization of grammatical strings over many trials in traditional AGL paradigms. Ziegler, Edelsbrunner, and Stern (2018) compared how learning through implicit vs. explicit tasks helped sixth-graders learn to solve algebraic addition and multiplication problems. Students in the implicit learning condition created and solved new problems to practice applying the underlying equation rules without being explicitly directed to attend to conceptual information. In contrast, students in the explicit learning condition were encouraged to extract principles and rules by studying example equations and generating verbal explanations. On three posttests administered over a 10-week period, students in the explicit learning condition showed improved problem-solving abilities and verbalization of concept knowledge compared to students in the implicit learning condition.

These two studies did not explore the synergy of implicit and explicit learning processes, but other work in the domain of mathematical cognition has explored the theory that different types of learning processes are intricately related, and that earlier acquisition of some knowledge may affect how later concepts are learned (e.g., Hazzan, 1999; Richland, Stigler, & Holyoak, 2012). For example, learning opportunities designed to improve procedural, nondeclarative knowledge can also benefit explicit, conceptual knowledge, and vice versa (Aleven & Koedinger, 2002; Rittle-Johnson, Schneider, & Star, 2015; Schneider & Stern, 2010). Lampinen and McLelland (2018) also explored how combining two presentations can benefit learning of a mathematical task. Adult participants were trained to perform a group operation on a cyclic group of order 6 in one of three presentational learning conditions: a modular arithmetic presentation, a visuospatial presentation, or a hybrid condition in which participants were presented with both presentations and told to integrate them. Participants' knowledge was

evaluated on a posttest of multiple problem types designed to test their understanding of the group operation and their ability to generalize this understanding to a cyclic group of order 9. Individual differences explained some of the variation in learning outcomes by condition, but showing participants both presentations and encouraging them to recognize the relationship between presentations improved posttest performance without requiring additional problem-solving time.

The Present Study. Previous work in both the general cognition and math learning literatures has demonstrated the value of using different learning processes or seeing two representations of a concept in order to bridge learners' understanding between nondeclarative, procedural knowledge and declarative, verbalizable knowledge. In the present study, we explore these findings in the context of a powerful learning device: gesture. Teachers routinely produce gestures as they instruct students in both one-on-one tutorials and the classroom (Goldin-Meadow, Kim, & Singer, 1999; Flevares & Perry, 2001). Students pay attention to their teachers' gestures, often gleaning important information from gesture that cannot be found anywhere in the teacher's speech (Singer & Goldin-Meadow, 2005). Compared to instruction with speech alone, gesture instruction can improve students' understanding of key math principles, such as *math equivalence* (the relation between two quantities that are the same), and these positive effects on learning persist over time (e.g., Cook, Mitchell, & Goldin-Meadow, 2008). Though benefits of gesture instruction on students' learning are well-established, it is unclear *why* gesture is such an effective learning tool. In the present study, we explore implicit and explicit learning through the lens of gesture by focusing on the similarities that learning through gesture shares with both learning processes, and in doing so, we aim to specify the mechanism by which gesture promotes learning.

As with implicit learning processes, observing gesture can lead to unconscious change in knowledge of a concept (Goldin-Meadow, Alibali & Church, 1993; Garber, Alibali & Goldin-Meadow, 1998; Broaders & Goldin-Meadow, 2010; Gurney, Pine & Wiseman, 2013; Kelly et al., 1999; Thompson, 1995; Kelly et al., 2004; Weismar & Hesketh, 1993). For example, Ellis Weismar and Hesketh (1993) taught children novel words through observed gestures or emphatic stress and found that children could not identify words they had learned through gesture, even though seeing the gestures had increased their subsequent comprehension of the novel words. Gesture is also an action, and the benefit of learning through gesture may partially arise from the lifelong experience of producing and seeing gesture (Ping, Goldin-Meadow, & Beilock, 2014). This salient feature of gesture is reminiscent of the connection between procedural non-declarative knowledge and the motor system (Seger, 1994; Shanks & St. John, 1994; Shanks, 2005). Though it shares many qualities with implicit learning, learning through gesture may also promote explicit learning processes. Unlike other types of movements, gesture is seamlessly integrated and routinely occurs with speech, a vehicle for declarative knowledge (Bernardis & Gentilucci, 2006; Church, Kelly, & Holcombe, 2014; Loehr, 2007; McNeill, 1992). Gesture and speech form one coherent system of communication, as these modalities are synchronized temporally (Nobe, 2000) and in the information they convey (McNeill, 1992). Furthermore, just as knowledge acquired through explicit learning processes can be flexibly applied to novel problems, learning through seeing gesture can also facilitate generalization ability (Cook, Duffy, & Fenn, 2013; Congdon et al., 2017; Wakefield et al., 2018).

We propose that gesture may interact with implicit and explicit learning processes to help learners transition from a state of non-declarative to declarative knowledge of a concept. In this way, gesture may help learners synergize their two knowledge states, much like explicit

knowledge can emerge from implicit knowledge. In exploring this hypothesis, we also aim to extend previous work on implicit and explicit instruction of math concepts in two key ways: 1) comparing implicit and explicit learning instructions (rather than learning tasks themselves) in a controlled experiment (thus expanding upon Ziegler et al., 2018); and 2) exploring how implicit and explicit learning instruction may work *synergistically* in a math learning context to benefit learning beyond one type of learning alone. By examining how gesture interacts with these questions, we also hope to clarify the mechanisms through which gesture benefits learning across the lifespan by extending previous work with gesture and math learning in children to adult learners.

In the present study, we manipulate adult participants' *intention to learn* the rules of a symbolic math equivalence system from Kaminski, Sloutsky, and Heckler (2008) by varying the *instructions* given to participants during the learning task. This study extends artificial grammar learning paradigms used in the learning literature to a novel task that can easily incorporate gesture. The symbolic math equivalence system we use is similar to both finite-state and biconditional artificial grammars used in previous work (e.g., Reber, 1976; Mathews et al., 1989). Like a finite-state grammar, the symbol system can create multiple similar exemplars (in this case, equations) that are either valid (correctly solved) or invalid (incorrectly solved). At the same time, these equations are constructed according to conditional rules (e.g., two symbols combine to form exactly one other symbol). We trained participants on equations in different learning conditions, in which we varied whether participants were told to simply remember the equations (implicit learning), to try to decipher the rules of the symbol system (explicit learning), or to follow implicit instructions for the first half of learning and to follow explicit instructions for the second half (implicit→explicit learning). After viewing videos of equations, participants

were then tested on both their ability to recognize equations seen during training, as well as their understanding of the underlying rule structure of the equations, and we examined whether learning persisted on a 24-hour followup test. This experiment thus tested how intention to learn, a distinguishing factor between implicit and explicit learning, functions in a math-learning paradigm. Additionally, by including an implicit→explicit condition, we explored how well participants learn the symbol system when they first establish an implicit knowledge base before engaging in explicit learning processes. Unlike a traditional AGL paradigm, gesture can be easily incorporated into our training paradigm of the symbolic system. Thus, half of the participants in the three learning conditions saw videos of a teacher performing a gesture that depicted a problem-solving strategy. This gesture was adapted from previous studies on the role of gesture in learning of math equivalence in children (e.g., Goldin-Meadow et al., 2009; Novack et al., 2014).

Given that our paradigm uses graphic shapes to makeup equations rather than traditional letter strings, and that it frames tasks in terms of “solving” equations, we did not have strong predictions about how implicit and explicit learning instructions would impact learning in our paradigm compared to in traditional AGL paradigms. There is reason to believe that the rule structure of the equations would be highly salient in our paradigm, as there were few rules to learn, these rules were simple two-symbol combinations, and participants saw multiple examples of equations of the same structure, giving them many chances to extract underlying rule structure. Furthermore, previous work has shown that participants can use graphic symbols as classification cues for associated grammatical letter strings even when given implicit instructions (e.g., Morgan, Meier, & Newport, 1987; Altmann et al., 1995). Our first research question thus

concerns whether implicit and explicit learning processes can be applied to a novel task not traditionally used in the implicit vs. explicit learning literature.

Our second research question asks how gesture will interact with implicit vs. explicit learning processes. One possibility is that by conveying non-declarative knowledge, gesture will simply benefit learning regardless of the context in which it is presented. Non-declarative knowledge can influence learning by guiding early processing of stimuli, rapidly providing new data to be processed by the declarative system (Musen & Treisman, 1990). By providing a visual representation of non-declarative knowledge, seeing gesture may make learners more likely to attend to task-relevant information. Furthermore, the gesture used in our paradigm specifically highlights the two-symbol combinations underlying the symbolic system, thus presenting learners with a non-declarative representation of the principles that all exemplars follow. In attending to this information, the non-declarative representations conveyed by gesture may provide learners with helpful information to construct a correct declarative representation of the symbolic system. If this is the case, we would expect to see an overall effect of gesture instruction on learning across conditions, particularly on items designed to test learners' understanding of the rules underlying the symbolic system. Gesture may also be particularly effective in the implicit→explicit condition, as these instructions further encourage participants to bridge their non-declarative knowledge with developing declarative knowledge.

If, however, gesture benefits learning by serving as a vehicle for implicit knowledge that also seamlessly integrates with explicit knowledge, gesture may better facilitate learning when it is presented in an explicit learning context than when it is presented in an implicit learning context. In this case, gesture would be grounded in an explicit context, allowing learners to benefit from the unspoken knowledge conveyed in gesture while consciously trying to learn how

to solve the problems. In contrast, when gesture is presented in an implicit learning context, learners may not fully engage with the cognitive processes necessary to take advantage of the information conveyed in gesture. In this case, we would expect to see an interaction of gesture instruction (gesture present vs. gesture absent) with learning type (implicit vs. explicit). It is unclear how gesture would function with the implicit→explicit condition in this case, but if learners are only able to benefit from gesture during the explicit half of training in this condition, we may expect to see performance somewhere in-between that of the implicit and explicit conditions. A final possibility is that gesture must be paired with meaningful co-speech in order to benefit learning. In other words, learners may struggle to glean non-declarative knowledge conveyed in gesture if this knowledge is not grounded in declarative knowledge conveyed in *speech* specifically. In this case, we would not expect to see any benefit of gesture across the different learning conditions, as the gestures in our paradigm were not paired with meaningful speech (see Method). Including the three gesture conditions thus allowed us to test whether gesture's effectiveness as a learning tool depends on the instructional context in which it is presented—specifically, whether learners benefit from gesture while engaging in either implicit or explicit learning processes, or a combination of both.

Method

We used a training followed by posttest design. 600 adult participants (*Mean age* = 37.7 yrs, *SD* = 11 yrs; 61.5% male, 38% female, 0.3% preferring not to say, 0.2% other gender) were recruited through Amazon Mechanical Turk and completed all components of the study via an online Qualtrics survey. The sample was racially and ethnically diverse (ethnicity: 71% identifying as not Hispanic, 25.8% Hispanic, 1% another ethnicity, 2.2% preferring not to say; race: 69.3% White, 17% Black, 7% Native American, 5.3% Asian, 0.8% more than one race,

0.3% preferring not to say, 0.2% other race). Participants were trained and tested on problems generated by a generic symbol system from Kaminski and colleagues (2008; see Table 1). Participants in all learning conditions were trained on the same equations in the same randomized order, but the instructions participants saw at the beginning and halfway through the training session varied by condition (see *Training*). The experiment thus had two independent variables, type of instruction (3 levels: implicit vs. explicit vs. implicit → explicit) and gesture (gesture present vs. gesture absent), resulting in a 3 x 2 design with six between-participants conditions: *implicit alone*, *implicit+gesture*, *explicit alone*, *explicit+gesture*, *implicit→explicit alone*, *implicit→explicit+gesture*. Following training, all participants completed an immediate posttest consisting of all equations seen during training, new problems of the same structure as those shown during training, and simple equations that tested participants' knowledge of the rule structure underlying all equations. Accuracy on the immediate posttest served as the dependent variable.

Training: All participants were trained on the same 24 symbolic equations, created according to the rules of the generic symbol system from Kaminski et al. (2008). These equations were presented in 4 training blocks, each in a fixed randomized order, for a total of 96 training trials. The instructions for both the implicit alone and implicit+gesture conditions were identical to each other, as were the instructions for both explicit learning conditions. After the second training block, participants in the implicit and explicit conditions were invited to take a short break, and prior to beginning the third block of training, they were presented with the same instructions they saw at the beginning of the study. For the implicit→explicit conditions, participants saw the implicit instructions prior to beginning the first 2 blocks of training and the

explicit instructions after the break, prior to the second 2 blocks of training. See Appendix A for full text of training instructions by condition.

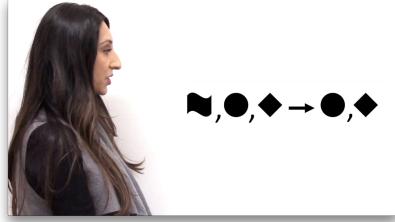
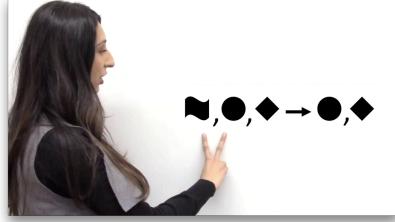
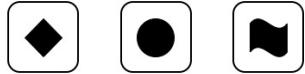
For each equation presented during training, participants saw a video of a teacher looking at the equation displayed on a whiteboard. The teacher's eye gaze and speech in the video were controlled for across conditions, with the experimenter saying the same neutral phrase for every equation: "Take a look at this problem." In the three gesture conditions, the teacher simultaneously performed a "grouping" gesture with this neutral speech (see Figure 1). The gesture was identical to that used in previous studies of mathematical equivalence in children (e.g., Goldin-Meadow et al., 2009; Novack et al., 2014): the teacher produced a V-handshape under two adjacent symbols on the left side of the equation and pointed to a symbol (or blank) on the right side of the equation to indicate how two symbols can be combined to make another symbol according to the rules of the system. After watching the video, participants in all conditions saw a new screen showing the same equation from the video, but with a symbol removed. Participants were asked to select the correct symbol to solve the equation. After they made their selection, participants were given feedback on their response and the correctly solved equation was shown for 5 seconds.

Given that there are three symbols in the system (diamond, circle, and squiggle), the order of the equations was designed so that throughout each block of the 24 equations, participants saw each symbol as the correct answer to an equation 8 times with no symbol being the correct answer for more than three problems in a row. For the gesture conditions, the teacher produced the V-point to the *first two* symbols on the left side of the equation for half of the equations and the *second two* symbols on the left side of the equation for half of the equations.

Table 1. The rules for combining symbols in the system from Kaminski et al. (2008).

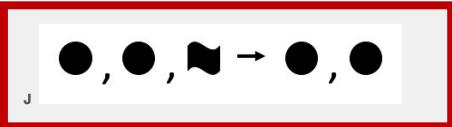
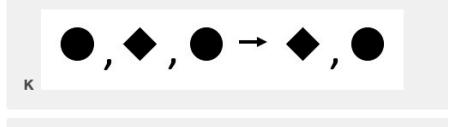
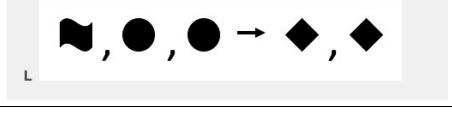
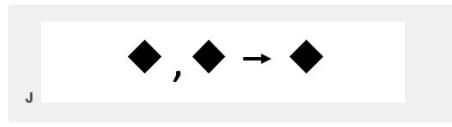
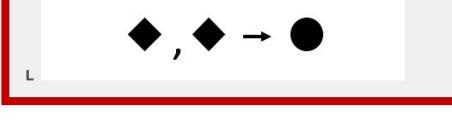
Rule 1: The order of two symbols does not change the result.
Rule 2: When a symbol combines with  , the result will always be the other symbol.
Rule 3:  ,  \rightarrow 
Rule 4:  ,  \rightarrow 
Rule 5:  ,  \rightarrow 
Rule 6: The result of more than two symbols does not depend on which two symbols combine first.

Table 2. Example of training trial.

1. Instructions (implicit or explicit) shown before training blocks 1 and 3; for the implicit→explicit conditions, implicit instructions were shown before training block 1, and explicit instructions were shown before training block 3	<p>Implicit: <i>Each video will show an equation. You should look at the equation and try to remember it.</i></p> <p>OR</p> <p>Explicit: <i>Each video will show an equation. You should look at the equation and try to remember it.</i></p> <p><i>The equations you will see follow a set of rules. The rules allow certain symbols to combine to make other symbols. Your task is to try to figure out what these rules are—that is, which symbols combine to make other symbols. You will be tested on your knowledge of these rules, so it is important that you try your best to learn them.</i></p>
2. See equation plus speech: “Take a look at this problem.”	<p>No Gesture</p>  <p>Gesture</p> 
3. Task (same for all conditions)	<p>What goes in the blank?</p> <p>♪, ●, ♦ → __, ♦</p> <p>Choose the best answer:</p> 
4. Feedback shown for 5 seconds (same for all conditions)	<p>Correct.</p> <p>♪, ●, ♦ → __, ♦</p> <p>OR</p> <p>Incorrect. The correct symbol is outlined in red.</p> <p>♪, ●, ♦ → __, ♦</p>  

Posttest and Follow-up: After training, all participants completed a 63-item posttest. Participants were introduced to the posttest by being told that they would see three equations at the same time, and that these equations would follow the same rules as the equations they had seen previously. Note that for participants in the implicit alone and implicit+gesture training conditions, this was the first mention of a rule structure underlying the equations. For each posttest item, participants were shown three equations at the same time: 1 correctly-solved and 2 incorrectly-solved. Participants were asked to choose the equation that they believed was solved correctly. Participants were unable to submit their choice until they had viewed the item for at least 3 seconds (this design element was added to discourage participants from rushing through or randomly guessing on the posttest items). Of the 63 posttest items, 24 items were the same equations seen during training, 21 items were novel equations of the same structure as those seen during training (three symbols on the left side of the equation, two symbols on the right side), and 18 items were two-symbol combination equations to assess participants' understanding of the rules underlying the symbolic system (9 unique problems, each shown twice). The posttest items were presented in a fixed, randomized order for all participants. The followup test given 24-hours after the immediate posttest was identical to the immediate posttest with the order of the problems presented in a fixed randomized order that differed from the order of the immediate posttest.

Table 3. Examples of posttest items with correct answers outlined in red.

Old Problems (seen during training) and New Problems (same format as trained problems)	Rule Problems
<p><i>Which equation is solved correctly?</i></p> <p>J </p> <p>K </p> <p>L </p>	<p><i>Which equation is solved correctly?</i></p> <p>J </p> <p>K </p> <p>L </p>

Questionnaire: Following completion of the follow-up posttest, participants answered a few short questions aimed at understanding how they processed the equations, whether or not they used gestures to help them (if applicable), and what rules (if any) they were able to decipher. These questions were designed to be open-ended and subjective in order to see whether participants could articulate any of the rules, and what they thought about the tasks in general. See Appendix B for full questionnaire.

Procedure: The experiment was conducted in two online sessions, lasting an average of 63 minutes for Session 1 and 23 minutes for Session 2. Participants were given informed consent prior to beginning Session 1, and they completed the entire study on their home computer. During the training, posttest, and followup phases of the study, participants pressed the “J,” “K”, or “L” keys, which corresponded to the answer choices for each item, to make their responses. Participants’ reaction time for each trial for each phase of the study was also recorded. Participants were paid \$4.50 for completing the first phase of the study.

About one-third of participants ($n = 208$) completed a followup posttest 24-hours after the immediate posttest (ns by condition: implicit alone = 37, implicit+gesture = 34, explicit alone = 31, explicit+gesture = 38, implicit→explicit alone = 26, implicit→explicit+gesture = 42). Amazon Mechanical Turk does not currently provide a researcher-friendly system for longitudinal studies (e.g., it is against MTurk policy to collect participant email addresses to send them a followup survey), and thus, attrition was high for the followup test. Participants were instructed on the first day of the study that they would receive a link to a followup survey via a message through MTurk's bonus system. Participants received the link to the survey in a message containing a \$0.01 bonus, and received an additional \$1.00 bonus if they completed the followup.

Results

Exclusions: Data were collected from 650 participants in total. Based on results from a pilot study prior to pre-registering the present study, we decided to exclude participants who did not perform significantly above chance on the training portion of the experiment, as such low accuracy likely indicated that the participant was not paying attention and/or did not understand the instructions of the task. There were 96 problems on the training, each with 3 choices (probability of answering correctly if randomly guessing = 0.33; probability of answering incorrectly if randomly guessing = 0.66). Expected chance performance was thus approximately 32/96 problems correct, and for a chi-square test where $p < 0.05$, the observed number correct would have to be 41 or more correct to be considered significantly above chance. Thus, we excluded participants who answered 40 or fewer training problems correct. 50 participants were excluded based on this criterion. Following this exclusion criteria, we continued data collection

until we collected data from 600 participants, 100 participants for each of the 6 training conditions.

In exploring differences in reaction time between conditions and controlling for reaction time in evaluating accuracy, we wanted to exclude outliers from the data. We first looked at trials on which participants had extremely high response times compared to the mean response time, as such response times likely indicated that participants took a short break from the study or got distracted on a particular trial (given that participants completed the study on their home computers rather than in a lab setting, this is a reasonable assumption). For the training and posttest phases of the study, the median reaction time was 4 seconds and the mean reaction time was 8.5 seconds ($SD = 68.4$ seconds). We removed trials from the data on which participants' reaction time was greater than three standard deviations from the mean reaction time, resulting in excluding 0.28% of trials. Additionally, we excluded reaction time for a small number of trials recorded as “0” seconds by Qualtrics, likely indicating that participants' response time for these trials was below the sensitivity threshold for Qualtrics' timing questions. This resulted in excluding an additional 0.1% of trials.

Analytical Approach: To explore whether training accuracy varied by type of instruction (implicit, explicit, or implicit→explicit), gesture (gesture absent or gesture present), or an interaction of the two, we conducted mixed-effects binomial logistic regression models for each phase of the study (training, posttest, and followup), with type of instruction (implicit; explicit; implicit→explicit), gesture (gesture present; gesture absent), reaction time, and interactions as main effects, and accuracy on each trial (0, 1) as the outcome variable. Participant and item were included as random effects in all models. For the posttest model, training accuracy and reaction time during training were also included as main effects to ensure that any effects of learning type

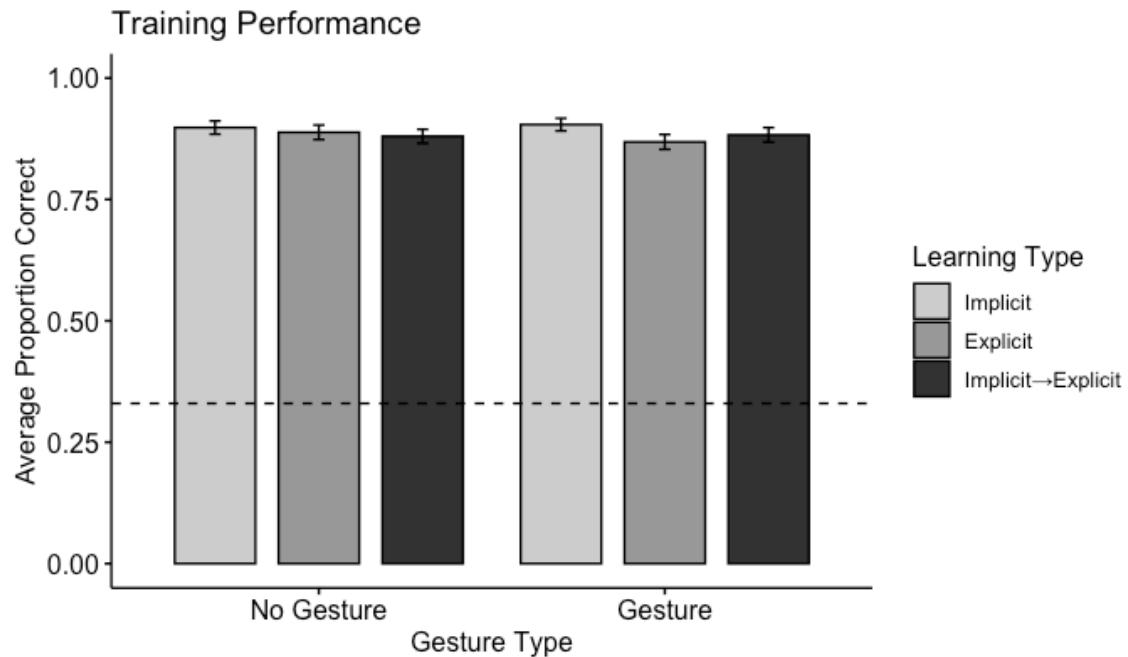
or gesture were not simply due to participants working through the training problems more slowly in some conditions versus others.

For each model we report, we present chi-square values that represent the main effects of each categorical variable of interest, beta values that reflect relative differences between levels of each categorical variable, and beta values that represent main effects of continuous covariates, such as reaction time.

Training – Accuracy: Accuracy on the training portion of the study was high for all learning types, with and without gesture. Average number of problems correct out of 96 training problems was as follows for the six training conditions: implicit alone – 90.2 (SD = 12.2), implicit+gesture – 90.7 (SD = 8.1), explicit alone – 91.6 (SD = 4.7), explicit+gesture – 85.3 (SD = 13.7), implicit→explicit alone – 89.9 (SD = 6.1), implicit→explicit+gesture – 86.3 (SD = 15.7).

A mixed-effects binomial logistic regression model with type of instruction (implicit, explicit, implicit→explicit), gesture (gesture present, gesture absent), and interactions as main effects; reaction time as a covariate; participant and item as random effects; and accuracy on each training trial (0, 1) as the outcome variable revealed no effects of learning type (χ^2 (2) = 3.3, p = 0.19), gesture (χ^2 (1) = 0.12, p = 0.73) nor an interaction (χ^2 (2) = 1.3, p = 0.53) on training accuracy. Reaction time was significantly negatively associated with accuracy (β = -0.16, z = -7.0 p < .00).

Figure 1. Training accuracy by learning type and gesture type. Proportion correct out of 96 items. Error bars represent standard error.



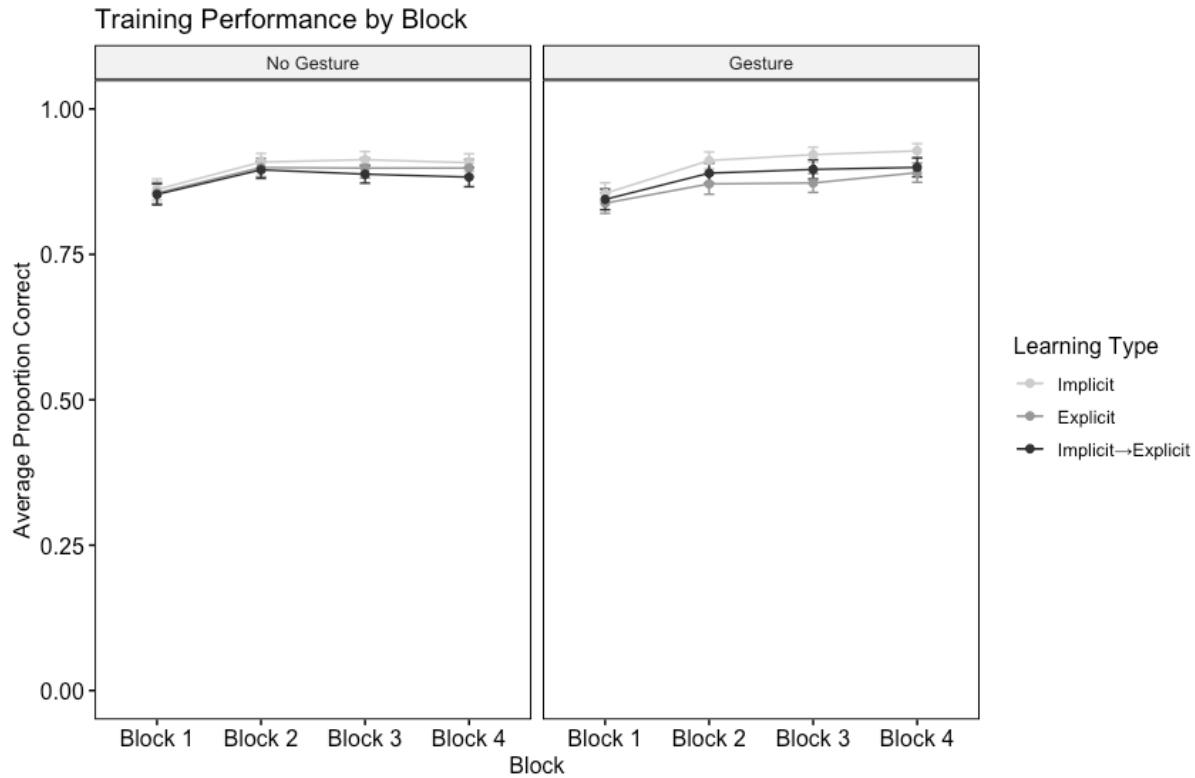
We also looked at training accuracy by Block. Figure 2 shows the average proportion correct for each training block (24 items per block) by learning type and gesture. To explore whether participants improved over the four training blocks by learning type or gesture, we ran a mixed-effects binomial logistic regression model with type of instruction (implicit, explicit, implicit→explicit), gesture (gesture present, gesture absent), training block (1, 2, 3, or 4), and interactions as main effects; participant as a random effect; and accuracy on each item (0, 1) as the outcome variable.

An analysis of variance of the model revealed a significant effect of block ($\chi^2 (3) = 307.4, p < 0.00$), an interaction of learning type and training block ($\chi^2 (6) = 16.1, p = 0.01$), and an interaction of gesture type and training block ($\chi^2 (3) = 12.7, p < 0.01$). Though participants' performance was high even on Block 1, participants in all learning conditions improved over the course of training. Participants performed better on Block 2 compared to Block 1 ($\beta = 0.70, z =$

6.9, $p < 0.000$). There was no significant improvement from Block 2 to 3 ($\beta = 0.16$, $z = 1.4$, $p = 0.15$). And while performance on Block 4 did not significantly differ from Block 3 ($\beta = 0.11$, $z = 0.96$, $p = 0.34$), performance on Block 4 was significantly better than Block 2 ($\beta = 0.27$, $z = 2.4$, $p < 0.02$), suggesting overall improvement from Block 1.

To explore the interactions between learning type and block and gesture type and block, we ran simpler mixed-effects binomial logistic regression models for each block to predict accuracy on each problem type by learning type and gesture for each block separately. Thus, we ran separate models of the same structure as the omnibus model for each block: Block 1, Block 2, Block 3, and Block 4. On Blocks 2 and 3, participants who received implicit learning instruction showed better performance than those who received explicit learning instruction (Block 2: $\beta = 0.55$, $z = 2.0$, $p < 0.05$; Block 3: $\beta = 0.58$, $z = 2.2$, $p < 0.03$). These results suggest that participants in the implicit conditions may have had an easier time with the training task as a result of their simpler instructions. On Block 3, participants who received implicit followed by explicit instruction and did *not* see gestures performed worse than participants who received this same instruction but did *see* gestures ($\beta = -0.74$, $z = -2.0$, $p = 0.04$). It is possible that seeing gesture helped participants transition more smoothly from implicit to explicit instruction than participants who did not see gesture.

Figure 2. Training accuracy across training blocks by learning type and gesture type. Proportion correct out of 24 items per block. Error bars represent standard error.



Training – Reaction Time: Figure 2 shows the average reaction time per trial for each training block (24 items per block) by learning type and gesture. To explore whether participants responded more quickly across the four training blocks by learning type or gesture, we ran a generalized linear model with type of instruction (implicit, explicit, implicit→explicit), gesture (gesture present, gesture absent), training block (1, 2, 3, or 4), and interactions as main effects to predict reaction time on each trial.

An analysis of variance of the generalized linear model revealed significant effects of learning type ($\chi^2 (2) = 43.9, p < 0.00$) and training block ($\chi^2 (3) = 52.8, p < 0.00$) on reaction time, and a trending interaction of learning type and training block ($\chi^2 (6) = 10.9, p = 0.09$). There was no main effect of gesture ($\chi^2 (2) = 0.59, p = 0.44$) on reaction time, and thus, we ran a subsequent model excluding this factor for simplicity. The explicit conditions were associated

with significantly higher reaction times than the implicit conditions (exp. vs. imp.: $\beta = 0.62$, $t = 2.8$, $p < 0.01$) and implicit→ explicit conditions (exp. vs. imp→exp.: $\beta = 1.1$, $t = 5.0$, $p < 0.00$), and the implicit conditions were associated with significantly higher reaction times than the implicit→explicit conditions (imp. vs. imp→exp.: $\beta = 0.51$, $t = 2.3$, $p = 0.02$). All learning type conditions performed significantly faster on block 4 compared to block 1 ($\beta = -0.7$, $t = -3.1$, $p < 0.01$). The implicit→explicit condition performed significantly slower on block 3 compared to all other training blocks ($\beta = 0.96$, $t = 3.0$, $p < 0.0$).

It is unsurprising that participants in the explicit conditions, who were told to learn the rules at the beginning of training, performed more slowly on training than participants in the other learning conditions. The difference in reaction time observed between the implicit and implicit→explicit conditions is somewhat surprising, as we expected participants in these conditions to look identical on reaction time for the first two blocks of training. However, participants in the implicit→explicit conditions do show an expected increase in reaction time for block 3, indicating that participants in these conditions followed the new instructions they received after block 2 to try to learn the rules on subsequent training trials. To ensure that reaction time did not account for differences in posttest accuracy, we included average reaction time per training trial as a covariate in our posttest model.

Figure 3. Training reaction time by learning type and gesture type. Error bars reflect standard error.

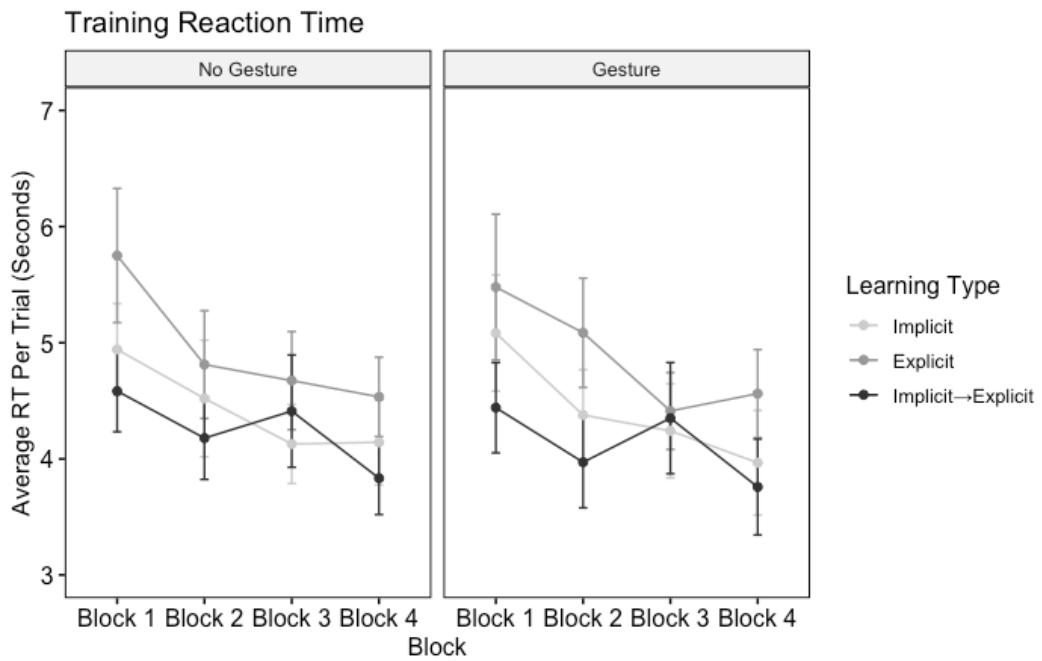
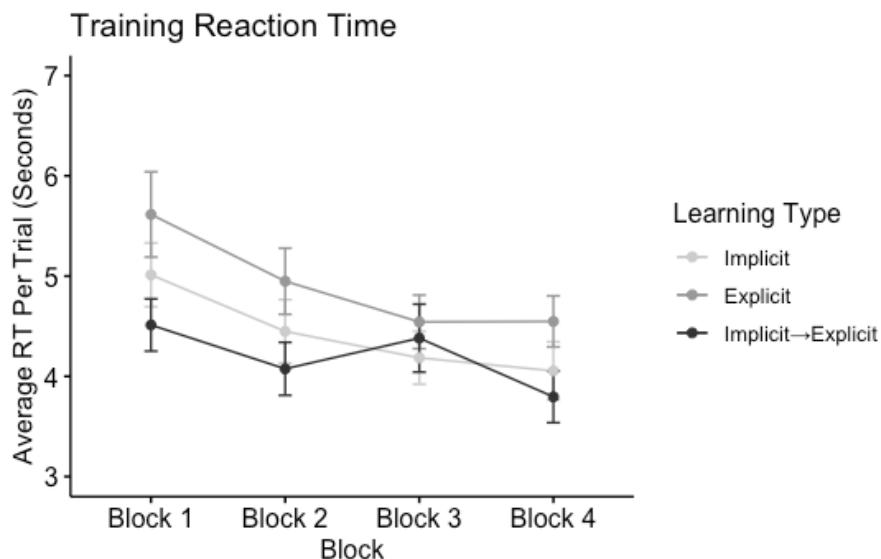


Figure 4. Training reaction time by learning type only (gesture conditions collapsed across learning type).



Posttest – Accuracy: Accuracy on the immediate posttest was low for all learning conditions. Average number of problems correct out of 63 posttest items was as follows for the six training conditions: implicit alone – 23 (SD = 5.8), implicit+gesture – 23.8 (SD = 8.5), explicit alone – 24.9 (SD = 8.6), explicit+gesture – 27.1 (SD = 12.6), implicit→explicit alone – 24.1 (SD = 8.9), and implicit→explicit+gesture – 26.2 (SD = 10.8). Despite low accuracy, participants in all conditions performed significantly above chance on posttest items: implicit alone – $t(99) = 3.8$, $p = < .00$; implicit+gesture – $t(99) = 3.6$, $p < .00$; explicit alone – $t(99) = 4.7$, $p < .00$; explicit+gesture – $t(99) = 5.0$, $p < .00$, implicit→explicit alone – $t(99) = 3.7$, $p < .00$; implicit→explicit+gesture – $t(99) = 5.0$, $p < .00$.

To examine accuracy by learning type and gesture, we ran a mixed-effects binomial logistic regression model with type of instruction (implicit, explicit, implicit→explicit), gesture (gesture present, gesture absent), problem type (old, new, rule), and interactions as main effects; reaction time, average reaction time per training trial, and training accuracy as covariates; participant and item as random effects; and accuracy on each training trial (0, 1) as the outcome variable. An analysis of variance of the model revealed significant effects of learning type ($\chi^2 (2) = 10.3$, $p = 0.01$), gesture type ($\chi^2 (1) = 6.1$, $p = 0.01$), and problem type ($\chi^2 (2) = 6.0$, $p < 0.05$), as well as interactions of learning type and problem type ($\chi^2 (4) = 10.2$, $p = 0.04$) and gesture type and problem type ($\chi^2 (2) = 9.5$, $p < 0.01$). Explicit instruction was associated with significantly better accuracy than implicit instruction (exp. vs. imp.: $\beta = 0.23$, $z = 2.9$, $p < 0.01$) and implicit→explicit instruction was associated with marginally better accuracy than implicit instruction (imp→exp vs. imp.: $\beta = 0.13$, $z = 1.7$, $p = 0.09$). Gesture was associated with better accuracy than non-gesture ($\beta = 0.14$, $z = 2.5$, $p = 0.01$). There were also significant effects of reaction time ($\beta = 0.03$, $z = 2.9$, $p < 0.01$), average reaction time per training trial ($\beta = -0.07$, $z =$

-2.6, $p < 0.01$), and training accuracy ($\beta = 0.14$, $z = 4.9$, $p < 0.00$), indicating that reaction time and training accuracy were positively associated with posttest accuracy, whereas higher average reaction time per training trial was associated with lower accuracy.

To further examine the interactions of learning type and gesture type with problem type, we used simpler models to predict accuracy on each problem type by learning type, gesture, and an interaction. Thus, we ran separate models of the same structure as the omnibus model for each problem type: old problems (problems seen during training), new problems (novel problems of the same structure as problems seen during training), and rule problems (problems designed to test participants understanding of the rules underlying the system). For old problems, there was a trending effect of gesture type ($\chi^2 (1) = 3.8$, $p = 0.05$) with the gesture conditions performing marginally better than the non-gesture conditions ($\beta = 0.12$, $z = 1.9$, $p = 0.05$). For new problems, there was a significant effect of learning type ($\chi^2 (2) = 8.5$, $p = 0.01$), with the explicit conditions performing significantly better than the implicit conditions ($\beta = 0.21$, $z = 2.9$, $p = 0.004$) and participants in the implicit → explicit conditions performing marginally better than the implicit conditions ($\beta = 0.12$, $z = 1.7$, $p = 0.08$); the explicit and implicit→explicit conditions did not significantly differ from each other ($\beta = 0.09$, $z = 1.2$, $p = 0.2$). For rule problems, there were significant effects of learning type ($\chi^2 (2) = 17.5$, $p < 0.001$) and gesture type ($\chi^2 (1) = 12.7$, $p < 0.001$). The explicit and implicit→explicit conditions performed significantly better than participants in the implicit conditions (explicit vs. implicit: $\beta = 0.32$, $z = 3.9$, $p < 0.000$; implicit→explicit vs. implicit: $\beta = 0.25$, $z = 3.1$, $p = 0.002$), but did not significantly differ from each other ($\beta = 0.07$, $z = 0.8$, $p = 0.4$). Participants in the gesture conditions performed better on rule problems than participants in the non-gesture conditions ($\beta = 0.24$, $z = 3.5$, $p < 0.001$).

Though there was no interaction of learning type and gesture type across problem types, we explored whether above-chance performance varied depending on the presence or absence of gesture during instruction. Mean accuracy and standard deviations as well as results of t-tests comparing accuracy to chance for each problem type by learning type and gesture are shown in Table 4. All learning types performed significantly above chance on old and new problems. Accuracy on rule problems differed significantly by learning type and gesture. The two implicit conditions did not perform reliably above chance on rule problems. However, in the explicit conditions and the implicit→explicit conditions, only participants who saw gesture during training performed significantly above chance on rule problems (though participants in the explicit alone condition performed marginally above chance). Welch's two-sample t-tests showed that participants in the explicit+gesture and implicit→explicit+gesture conditions performed significantly better than their non-gesture counterparts: (exp+gesture vs. exp.: $t(179.04) = 2.5$, $p = 0.01$; imp→exp.+gesture vs. imp→exp.: $t(192.6) = 2.2$, $p = 0.03$).

Figure 5. Posttest accuracy on all problem types combines by learning type and gesture type
Note: models were run at the trial level; here, we convey accuracy at the mean level by proportion correct out of 63 items (trials). Error bars represent standard error.

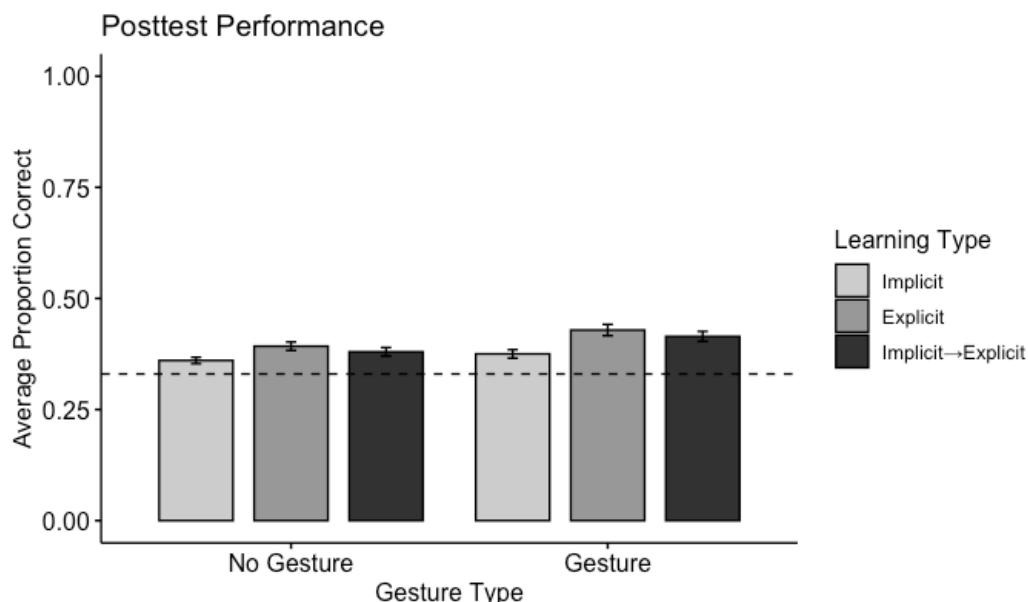


Figure 6. Posttest proportion correct on each problem type by learning type and gesture type. Error bars reflect standard error.

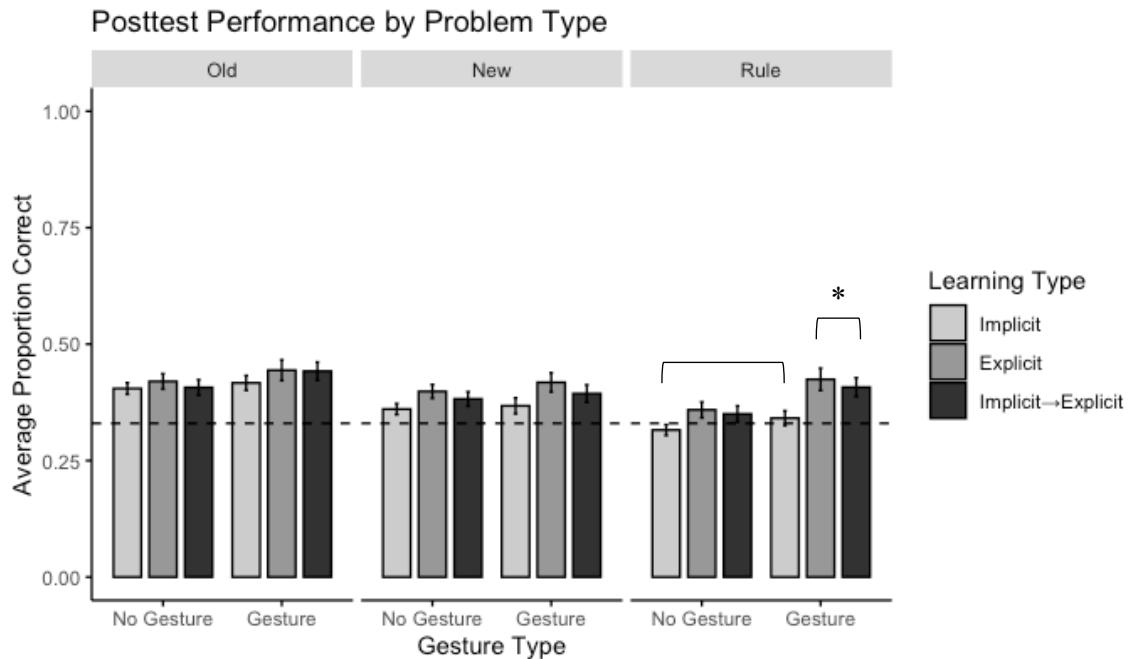


Table 4. Mean and standard deviation of posttest accuracy by problem type by learning type and gesture type; t-tests comparing accuracy on each problem type to chance performance (0.33 of problems correct; * indicate performance significantly above chance).

	Learning Type	“Old” Problems (24 Items)		New Problems (21 Items)		Rule Problems (18 Items)	
		No Gesture	Gesture	No Gesture	Gesture	No Gesture	Gesture
Means and Standard Deviations	Implicit	9.7 (2.9)	10.0 (3.8)	7.6 (2.5)	7.7 (3.5)	5.7 (2.1)	6.1 (2.8)
	Explicit	10.1 (4.0)	10.7 (5.3)	8.4 (3.1)	8.8 (4.3)	6.5 (3.1)	7.6 (4.3)
	Implicit→Explicit	9.8 (4.0)	10.6 (4.7)	8.0 (3.3)	8.3 (3.9)	6.3 (3.1)	7.3 (3.6)
t-tests comparing to chance performance	Implicit	t(99) = 6.1, p < .00*	t(99) = 5.4, p < .00*	t(99) = 2.6, p = .01*	t(99) = 2.2, p = .03*	t(99) = -1.1, p = .30	t(99) = 0.7, p = 0.5
	Explicit	t(99) = 5.4, p < .00*	t(99) = 5.2, p < .00*	t(99) = 4.3, p < .00*	t(99) = 4.3, p < .00*	t(99) = 1.7, p = .09	t(99) = 3.9, p < .00*
	Implicit→Explicit	t(99) = 4.7, p < .00*	t(99) = 5.4, p < .00*	t(99) = 2.6, p = .01*	t(99) = 2.2, p = .03*	t(99) = 1.2, p = .24	t(99) = 3.8, p < .00*

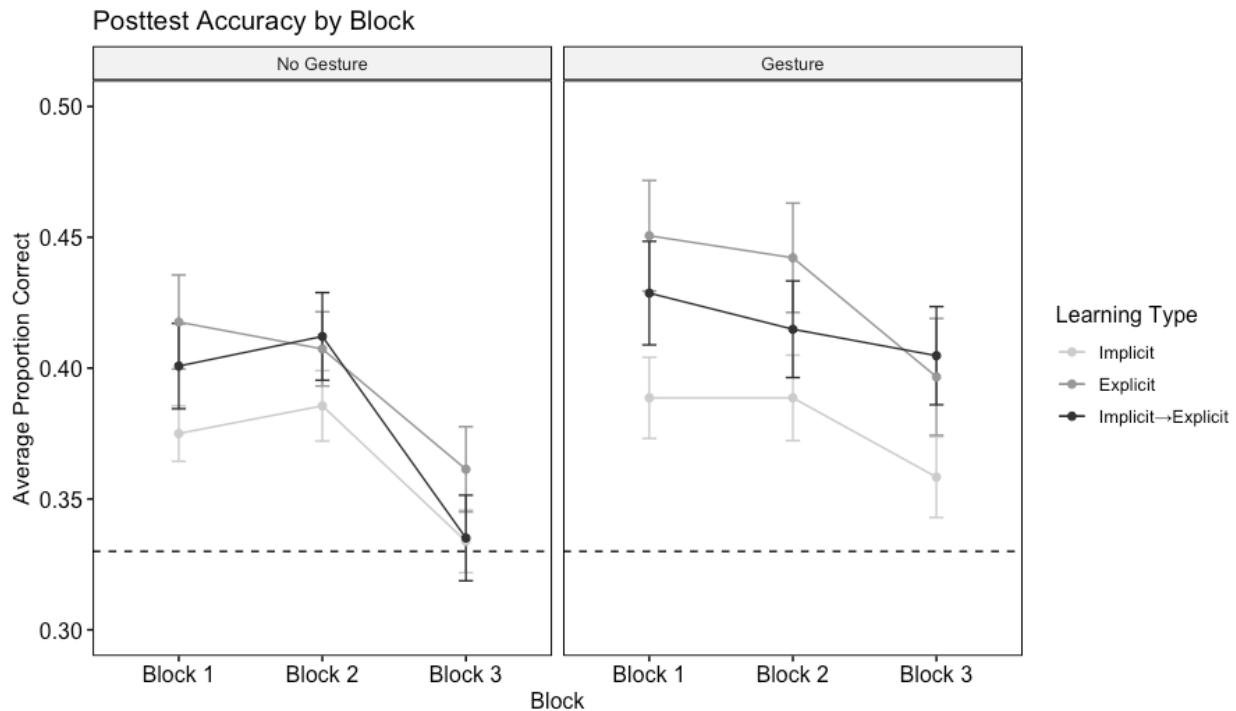
We also looked at posttest accuracy by Block. Figure 7 shows the average proportion correct for each posttest block (21 items per block) by learning type and gesture. To explore whether participants improved over the three posttest blocks by learning type or gesture, we ran a mixed-effects binomial logistic regression model with type of instruction (implicit, explicit, implicit→explicit), gesture (gesture present, gesture absent), training block (1, 2, 3), and interactions as main effects; reaction time, average reaction time per training trial, and training accuracy as covariates; participant and item as random effects; and accuracy on each posttest trial (0, 1) as the outcome variable.

An analysis of variance of the model revealed significant effects of learning type ($\chi^2 (2) = 10.3, p < 0.01$), gesture type ($\chi^2 (1) = 6.1, p = 0.01$), and block ($\chi^2 (2) = 76.2, p < 0.000$); and an interaction of gesture type and block ($\chi^2 (2) = 6.3, p = 0.04$). There were also significant effects of reaction time ($\beta = 0.03, z = 2.5, p = 0.01$), average reaction time per training trial ($\beta = -0.07, z = -2.6, p < 0.01$), and training accuracy ($\beta = 0.14, z = 4.9, p < 0.000$), indicating that reaction time and training accuracy were positively associated with posttest accuracy, whereas higher average reaction time per training trial was associated with lower accuracy. Participants' performance across conditions did not significantly differ from Block 1 to Block 2 ($\beta = -0.01, z = -0.1, p = 0.9$), but worsened from Block 2 to Block 3 ($\beta = -0.24, z = -3.5, p < 0.00$).

To explore the interaction between gesture type and block, we ran simpler mixed-effects binomial logistic regression models for each block to predict accuracy on each problem type by gesture type for each block separately. On Block 1, participants who saw gestures during training performed marginally better than participants who did not see gestures during training ($\beta = 0.12, z = 2.0, p = 0.05$). On Block 3, participants who saw gestures during training performed significantly better than participants who did not see gestures during training ($\beta = 0.21, z = 3.2, p$

$= 0.001$). Thus, seeing gesture during training helped participants most on the worst-performing block of the posttest.

Figure 7. Posttest proportion correct on each block (out of 21 problems) by learning type and gesture type. Error bars reflect standard error.



Posttest – Reaction Time: Figure 8 shows the average reaction time per item per problem type (24 old items, 21 new items, 18 rule items) broken down by learning type and gesture. To explore whether participants responded more quickly on some problem types at posttest as a function of their training experience, we ran a generalized linear model with type of instruction (implicit, explicit, implicit→explicit), gesture (gesture present, gesture absent), and problem type (old, new, and rule), and interactions as main effects to predict reaction time.

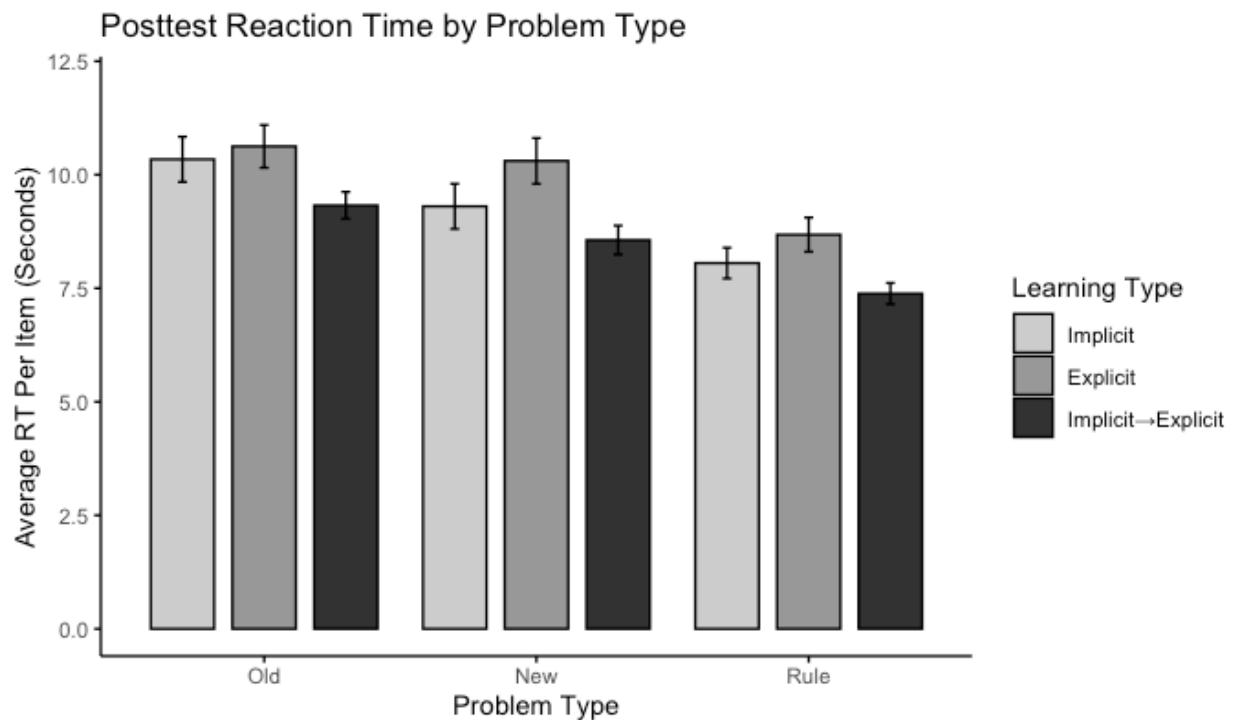
An analysis of variance of the generalized linear model revealed significant effects of learning type ($\chi^2 (2) = 86.8, p < 0.00$) and problem type ($\chi^2 (3) = 172.0, p < 0.00$), and no effect of gesture ($\chi^2 (2) = 1.3, p = 0.25$). Participants were slower on old problems compared to the

other two problem types (old vs. new: $\beta = 0.71$, $t = 4.7$, $p < 0.00$; old vs. rule: $\beta = 2.0$, $t = 13.0$, $p < 0.00$), and slower on new problems compared to rule problems ($\beta = 1.3$, $t = 8.3$, $p < 0.00$). This is somewhat surprising, as participants may have been expected to perform more quickly on items containing equations they had seen during training, and more slowly on novel equations and rule equations.

We further examined the main effect of learning type by using simpler generalized linear models to predict reaction time on each problem type by learning type. The explicit conditions performed significantly slower than the implicit→explicit conditions on all problem types (*old*: $\beta = 1.3$, $t = 4.5$, $p < 0.00$; *new*: $\beta = 1.7$, $t = 6.4$, $p < 0.00$; *rule*: $\beta = 1.3$, $t = 5.9$, $p < 0.00$), and slower than the implicit conditions on new and rule problems (*new*: $\beta = 1.0$, $t = 3.7$, $p < 0.00$; *rule*: $\beta = 0.64$, $t = 2.9$, $p < 0.00$). The implicit conditions were slower than the implicit→explicit conditions on all problem types (*old*: $\beta = 0.97$, $t = 3.4$, $p < 0.00$; *new*: $\beta = 0.73$, $t = 2.7$, $p < 0.01$; $\beta = \text{rule}$: 0.68 , $t = 3.0$, $p < 0.01$). Participants in the implicit→explicit conditions performed fastest on training, so it is possible that the reaction time advantage of this condition at posttest reflects a general trend of this condition performing more quickly on the study overall. Furthermore, it is unsurprising that participants in the implicit conditions, who were told about rules for the first time at the beginning of the posttest, would perform more slowly on these problem types compared to the implicit→explicit condition, which had already been made aware of rules. Still, the finding that the implicit→explicit conditions performed more quickly and more accurately on rule problems than their implicit and explicit counterparts, particularly in the gesture conditions, suggests that this training condition may have promoted faster processing of equations during posttest. Importantly, these differences by learning type on reaction time do not

fully account for differences in accuracy, as reaction time was included as a covariate in all of our accuracy models.

Figure 8. Posttest RT by problem type and learning type. Error bars reflect standard error.

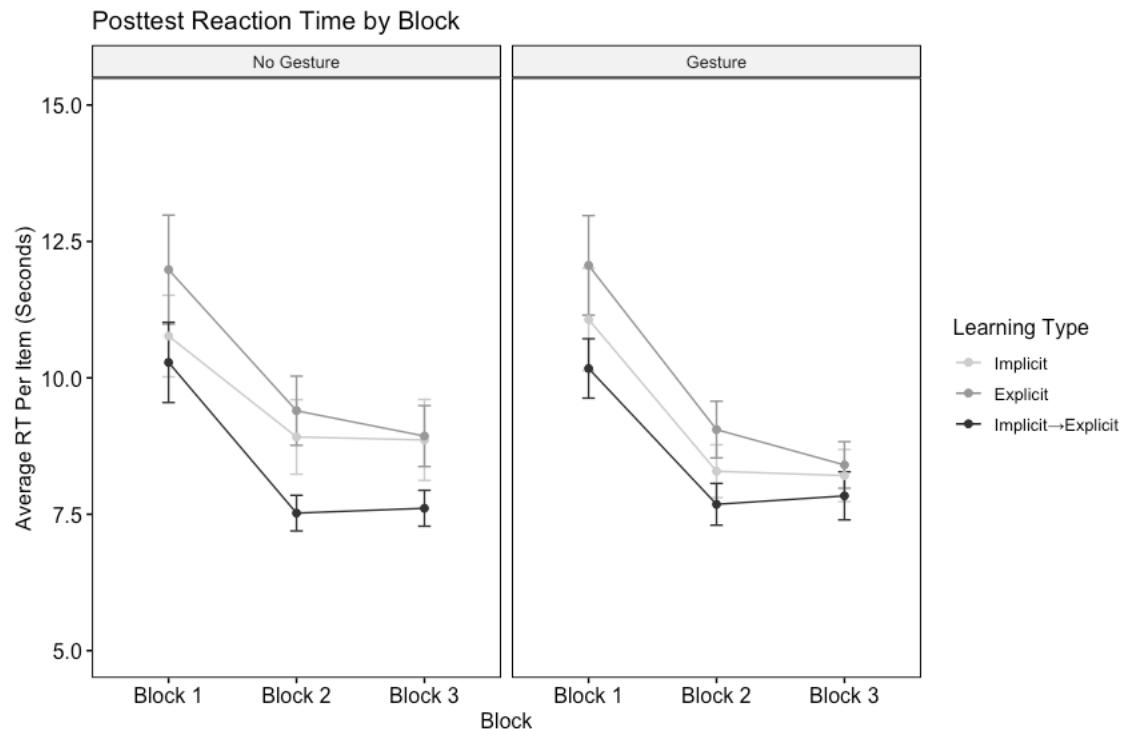


We also examined posttest reaction time over the course of the posttest. Figure 9 shows the average reaction time per trial for each posttest block (21 items per block) by learning type and gesture. To explore whether participants responded more quickly across the three training blocks by learning type or gesture, we ran a generalized linear model with type of instruction (implicit, explicit, implicit→explicit), gesture (gesture present, gesture absent), posttest block (1, 2, or 3), and interactions as main effects to predict reaction time on each trial.

An analysis of variance of the generalized linear model revealed significant effects of learning type ($\chi^2(2) = 87.6, p < 0.00$) and block ($\chi^2(2) = 390.4, p < 0.00$) on reaction time.

Across conditions, participants performed more quickly on Blocks 2 and 3 compared to Block 1 (Block 2 vs. Block 1: $\beta = -3.0$, $t = -7.8$, $p < 0.000$; Block 3 vs. Block 1: -3.6 , $t = -9.5$, $p < 0.000$).

Figure 9. Average reaction time per item on each posttest block by learning type and gesture type.



Followup Test: Attrition was high for the 24-hour followup test, and thus, we did not run statistical analyses on the followup data, but present the means and trends. Average number of problems correct out of 63 followup items was as follows for the six training conditions: implicit alone – 23.8 ($SD = 7.8$), implicit+gesture – 26.9 ($SD = 10.9$), explicit alone – 28.1 ($SD = 11.4$), explicit+gesture – 32.1 ($SD = 15.7$), implicit→explicit alone – 28.3 ($SD = 13.1$), and implicit→explicit+gesture – 28.0 ($SD = 12.4$). In comparing posttest performance to followup performance for participants who completed both assessments, the explicit+gesture and implicit → explicit gesture conditions showed patterns of maintained or even slightly improved performance from day 1 to day 2 across problem types, whereas the explicit and implicit →

explicit alone conditions showed patterns of worse performance from day 1 to day 2. The implicit+gesture condition showed a pattern of improved performance on old problems from day 1 to day 2, suggesting that gesture may have helped participants in this condition remember the items they were trained on. Overall, these patterns are in-line with previous work demonstrating that gesture's impact on learning may emerge and persist over time.

Figure 10. Posttest versus followup accuracy on each problem type by learning type and gesture for 208 participants who completed both assessments.

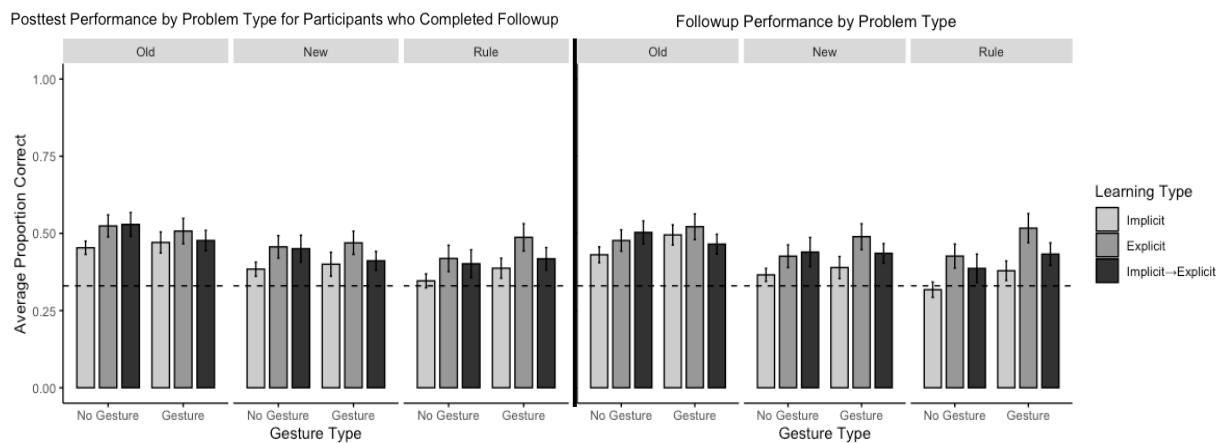


Table 5. Followup accuracy on each problem type by learning type and gesture type – Mean (Standard Deviation).

Learning Type	“Old” Problems (24 Items)		New Problems (21 Items)		Rule Problems (18 Items)	
	No Gesture	Gesture	No Gesture	Gesture	No Gesture	Gesture
Implicit	10.3 (3.7)	11.9 (4.6)	7.7 (2.7)	8.2 (4.4)	5.7 (2.7)	6.8 (3.3)
Explicit	11.5 (4.6)	12.6 (6.1)	9.0 (4.3)	10.2 (5.4)	7.7 (3.9)	9.3 (5.2)
Implicit→Explicit	12.1 (4.6)	11.1 (5.0)	9.2 (5.1)	9.1 (4.3)	7.0 (4.2)	7.8 (4.2)

Questionnaire: Only participants who completed the followup also completed the questionnaire. This was by design, as we did not want to question the participants about their knowledge of the rules prior to having them complete the followup test, as this could potentially

bias their performance on the followup. The final question of the questionnaire explicitly asked participants whether they had learned any rules for solving the equations, “*If you were told to learn the rules for solving the equations at the beginning of the experiment yesterday, or you think you may have an idea of what the rules are, please list them here. It's okay if you are uncertain about your answers.*” Only 29 out of the 208 followup participants correctly identified one or more of the rules underlying the system, and the distribution of rule-identifiers varied by condition (5 in explicit alone, 9 in explicit+gesture, 3 in implicit→explicit alone, 5 in implicit→explicit+gesture, 3 in implicit alone, and 4 in implicit+gesture). Of the 18 participants in the gesture conditions who correctly identified one or more rules, 9 of these participants indicated on an additional question that they had used the gestures to help them with the task.

Discussion

Participants in all three learning conditions (regardless of the presence of gesture) were able to identify correctly-solved trained and novel equations above chance levels at posttest. This result indicates that participants can learn to apply mathematical knowledge gained through exposure to multiple valid exemplars to identify novel valid exemplars of the same structure, regardless of whether this knowledge is acquired through implicit or explicit processes, or a combination of both processes. This result replicates previous findings from artificial grammar learning tasks showing that participants can correctly identify novel grammatical strings after exposure to valid strings, regardless of whether participants are told to simply memorize strings (e.g., Reber, 1976; Reber & Lewis, 1977; Brooks, 1978), or to explicitly learn the rules underlying the strings (e.g., Reber et al., 1980; Mathews et al., 1989). Despite the finding that implicit instruction led to some learning at posttest, the presence of at least some explicit instruction led to superior learning outcomes, as both the explicit and implicit→explicit

conditions performed better across all problem types than the implicit condition, and significantly better on items testing rule knowledge. Overall, our results demonstrate that a rule-based math concept may be learned through similar implicit and explicit learning processes as those engaged during artificial grammar learning tasks, but that some degree of explicit learning leads to more robust learning.

In exploring how gesture interacts with implicit and explicit learning processes, our results suggest that gesture is most beneficial for learning when it is presented with at least some explicit instruction. When we explored the observed interaction between gesture type and problem type, we found that gesture only benefited accuracy on rule problems when it was presented in either an explicit or implicit→explicit learning context. Despite above-chance performance on trained and novel problems at posttest, participants in the implicit conditions (even the implicit+gesture condition) did not perform above chance levels on identifying valid equations depicting rules underlying the mathematical system. However, the observed pattern of improved rule problem accuracy on the followup test by a small number of participants in the implicit+gesture condition indicates that perhaps the benefits of gesture in an implicit context may emerge over time. This is in line with previous work showing that the beneficial effects of gesture on learning may not emerge until at least 24-hours after training on a concept (Cook, Duffy, & Fenn; 2013; Wakefield et al., 2019; Congdon et al., 2017), just as implicit learning can occur gradually over time with repeated practice (Eichenbaum & Cohen, 2001; Henke, 2010).

Our findings have implications for the cognitive and gesture literatures, as well as for the educational literature, particularly for examining how multiple learning processes and exemplars may benefit learning of math concepts. Future work could further specify *why* our implicit paradigm led to transfer on novel problems, but not rule problems, and whether rule learning is

possible through more gradual implicit learning over time. Furthermore, many researchers have noted that “learning” within the realm of mathematical cognition can refer to factors such as conformance to rules when solving problems procedurally, explicit awareness of rules, ability to transfer rules to an analogous situation, or some combination of these abilities, and that the inferences we make about learners’ level of understanding can depend upon which of these features we use to evaluate learning (e.g., Bisanz & LeFevre, 1992). For instance, Greeno and Riley (1987) have shown that students can possess the ability to execute a procedure without having the ability to articulate the rules that the procedure follows. Perhaps implicit learning of a mathematical concept, like for other domains of implicit learning, can lead to transfer to an analogous situation (i.e., novel problems of a similar format) but not to rule conformance.

Our result that the implicit→explicit condition led to equally good learning compared to the explicit condition replicates previous work in the implicit and explicit learning literatures (e.g., Mathews et al., 1989) and math learning literatures showing that engaging in multiple learning processes can benefit learning (e.g., Schwartz and Goldstone, 2015; Rau, 2016; Lampinen & McLelland, 2018). It would be useful to further explore *why* the implicit→explicit condition led to positive learning outcomes, and whether this effect is due to individual differences (as observed by Lampinen & McLelland, 2018) or to an overall benefit of creating a nondeclarative knowledge base prior to engaging in explicit learning processes. This question could be tested by using our paradigm in conjunction with other cognitive measures, such as working memory and executive function, and by exploring the precise timing at which participants who receive implicit→explicit instruction acquire verbalizable, declarative knowledge compared to participants who receive only explicit instruction. For example, researchers could train participants in an explicit alone condition and test them at multiple time

points to evaluate when they demonstrate full knowledge of the rules of the symbolic system—that is, identify at what time point declarative knowledge “emerges.” This procedure could be replicated with a group who receives implicit→explicit training. If participants who first receive implicit instruction demonstrate explicit knowledge after less explicit training than participants in the explicit alone condition, this finding could indicate that the implicit training created a nondeclarative knowledge base from which participants were able to explicate declarative knowledge with minimal explicit training. In contrast, if participants require the same amount of explicit training as participants in the explicit alone condition to demonstrate knowledge of the rules, this may suggest that the implicit learning did not contribute to the acquisition of declarative knowledge (though our posttest reaction time data for the implicit→explicit condition suggests that implicit instruction prior to explicit instruction may still lead to faster *processing* of novel exemplars).

The posttest results for the gesture conditions offer insight into how gesture may benefit learning. Though a main effect of gesture was observed on posttest, this effect was primarily driven by performance by the explicit+gesture and implicit→explicit+gesture conditions on problems that tested participants’ understanding of the rules underlying the symbolic system, which all of the equations they viewed during training adhered to. The fact that gesture training in conjunction with some explicit instruction was beneficial specifically for these types of problems suggests that gesture may be particularly helpful for learning when participants *intend* to learn a concept. Some participants in all of the gesture conditions who completed the followup test were able to verbalize at least one rule, but not all of these participants who were able to verbalize rule knowledge reported finding the gestures helpful in acquiring this knowledge. This finding supports evidence from previous work that gesture’s effects on learning are often not

consciously accessible or verbalized (Goldin-Meadow, Alibali & Church, 1993; Garber, Alibali & Goldin-Meadow, 1998). Thus, gesture may be processed unconsciously by some learners, but it appears to be most beneficial when the learner is explicitly told to learn a concept.

The gestures used in our study were specifically intended to highlight rule structure. Future research should address whether different types of gesture can similarly benefit learning—that is, if the type of information conveyed in gesture impacts gesture’s effect on learning in an explicit learning context. Additionally, our finding that gesture benefited learning while being paired with meaningless speech warrants more investigation. Though gesture and speech form one coherent system for communicating information (e.g., McNeill, 1992), our results suggest that learners can glean meaningful information from gesture even when the gesture is not paired with informative speech. This finding supports our hypothesis that gesture may serve as a vehicle for non-declarative, procedural knowledge. Future work can further address how gesture may function with declarative knowledge conveyed in speech to help learners transition from a state of non-declarative knowledge to one of declarative, verbalizable knowledge.

CHAPTER THREE: DOES GESTURE DEPEND ON SPEECH TO BENEFIT LEARNING?

Introduction

Learners benefit from observing teachers who gesture while explaining a concept.

Children show improved knowledge of a variety of concepts after viewing a lesson where a teacher speaks while gesturing as opposed to speaking alone (e.g., Singer & Goldin-Meadow, 2005; Ping & Goldin-Meadow, 2008). Gesture has also been shown to impact how adults comprehend a speaker's speech (Kelly, Barr, Church, & Lynch, 1999; Thompson, 1995; Ping, Goldin-Meadow, & Beilock, 2014) and even learn new information (Kelly, McDevitt, & Esch, 2009; Cook & Fenn, in prep.). Moreover, the effects of learning through seeing gesture can emerge and persist over time (Cook, Duffy, & Fenn, 2013; Congdon et al., 2017; Wakefield et al., 2019; Kelly, McDevitt, & Esch, 2009; Cook & Fenn, in prep.).

Several studies have attempted to specify the mechanism through which gesture benefits learning. Evidence from these studies suggest that gesture can help learners follow along more closely with a teacher's speech (Wakefield et al., 2018) and activate a learner's own motor system (Ping et al., 2014), perhaps leading to rich sensorimotor representations in the brain of acquired knowledge. Another proposed mechanism that has not been fully explored is that gesture may convey implicit knowledge to a learner. Information learned through gesture shares many characteristics with information that is learned implicitly. Despite its profound ability to change learners' knowledge of a concept, gesture's influence on learners' behavior is often not consciously accessible or verbalized (Goldin-Meadow, Alibali & Church, 1993; Garber, Alibali & Goldin-Meadow, 1998). Observing gesture can change a learner's knowledge even if the learner is unable to verbally report that this knowledge came from gesture (Broaders & Goldin-Meadow, 2010; Gurney, Pine & Wiseman, 2013; Kelly et al., 1999; Thompson, 1995; Kelly et

al., 2004). Furthermore, gesture is an action, and the benefit of learning through gesture may partially arise from the lifelong experience of producing and seeing gesture (Ping et al., 2014). This salient feature of gesture is reminiscent of the connection between procedural non-declarative knowledge and the motor system (for reviews, see Seger, 1994; Shanks & St. John, 1994; Shanks, 2005).

Though previous work suggests that gestural information may be learned implicitly, gesture also has a unique relationship with speech, the gold standard for declarative knowledge. Declarative knowledge, acquired through explicit learning processes, is easily articulated by a learner. Unlike other types of movements, gesture forms a coherent system of communication with the speech it accompanies (Bernardis & Gentilucci, 2006; Church, Kelly, Holcombe, 2014; Loehr, 2007; McNeill, 1992; Nobe, 2000). As a result of its privileged relationship with speech, gesture may promote learning by functioning as a vehicle for unspoken nondeclarative knowledge that is seamlessly paired with speech, a vehicle for declarative knowledge. Thus, perhaps gesture is most effective for learning when it is grounded in meaningful speech.

In the present study, we ask whether unspoken, non-declarative knowledge conveyed in gesture is optimally useful for learning when it is paired with declarative knowledge—here defined as meaningful speech. To explore this question, we train participants to solve novel symbolic math equivalence problems created from a system developed by Kaminski and colleagues (2008) in one of three learning conditions: *declarative knowledge alone* (learners see a teacher explain problems using an equalizer strategy in *speech alone*), *non-declarative knowledge alone* (learners see a teacher produce an equalizer gesture with neutral speech: “Take a look at this problem”), or *simultaneous declarative and non-declarative knowledge* (learners see a teacher produce an equalizer strategy in speech with a matching gesture). Thus, we test 1)

whether gesture is more effective for learning when it reiterates spoken, declarative knowledge, as opposed to just providing unspoken, declarative knowledge, and 2) whether the non-declarative knowledge provided by instruction with gesture alone is more, less, or equally effective for learning than speech alone instruction.

If gesture benefits learning by depicting non-declarative knowledge that is also seamlessly integrated with speech, a vehicle for declarative knowledge, then gesture may be less beneficial for learning when it is presented with vacuous speech. When gesture is produced simultaneously with speech, it can represent a visuospatial representation of a concept at the same time as a verbal representation of the concept (Kendon, 1980; McNeill, 1992; Morrel-Samuels & Krauss, 1992). Indeed, instruction with gesture that reinforces information conveyed with speech has been shown to promote better learning than the same spoken instruction presented without gesture (Perry, Berch, & Singleton, 1995; Valenzano, Alibali, & Klatzky, 2003; Wakefield et al., 2018b; Congdon et al., 2017).

However, there is also evidence to suggest that gesture may promote learning in some contexts when presented on its own, in the absence of relevant speech instruction. In the current experiment, participants were informed of the rules of the symbolic math equivalence system prior to watching instructional videos with or without gesture, and thus were provided with some declarative knowledge on which to build. Work in social psychology has shown that individuals who have discrepancies between their implicit and explicit attitudes or beliefs are motivated to consider new relevant information in order to resolve these discrepancies (Brinol, Petty & Wheeler, 2006; Rydell, McConnell, & Mackie, 2008). Based on this account, seeing gestures that convey new information, or convey information in a novel format, could create a type of cognitive dissonance that encourages learners to engage in increased processing of information

relevant to the task to be learned. In other words, the non-declarative knowledge embodied in gesture may not itself change thinking but may, in conjunction with already held declarative knowledge, encourage learners to seek new information that will subsequently change thinking. Approaching gesture instruction with intact declarative knowledge may thus help participants glean non-declarative knowledge from the gestures they see (whether or not these gestures are paired with meaningful speech), and subsequently update their declarative knowledge of the symbolic math equivalence system. If this is the case, gesture alone instruction may promote learning as well as speech + gesture instruction, or at least better than instruction with speech alone. In fact, research from math equivalence learning in children suggests that it is possible to benefit from gesture instruction *alone* (without accompanying speech) when learners are asked to *do* the gestures themselves. Cook, Mitchell, & Goldin-Meadow (2008) found that children learned to solve math equivalence problems equally well when they were instructed to either perform simultaneous speech and gesture, or to just perform a gesture, as measured by performance on a delayed posttest following training. Both gesture conditions performed better than a speech alone condition in which children were taught to say a strategy in speech, but not to gesture. Furthermore, evidence from Chapter 2 of this dissertation suggests that *observed* gesture can promote some learning in an implicit learning context even when gesture is paired with meaningless speech.

Method

We used a pretest-training-posttest design. 309 adult participants (*Mean age* = 37.1 yrs, SD = 11.5 yrs; 61% male, 39% female) were recruited through Amazon Mechanical Turk and completed all components of the study via an online Qualtrics survey. The sample was racially and ethnically diverse (ethnicity: 61% identifying as not Hispanic, 35% Hispanic, 3% another

ethnicity, <2% preferring not to say; race: 74% White, 14% Black, 7% Native American, 3% Asian, 1% more than one race, <1% preferring not to say). All participants were introduced to the algebraic rules of the generic symbol system from Kaminski and colleagues (2008) and completed a pretest of problems generated according to the rules of the system (note that the symbols used differ slightly from those used in Study 1). Participants were then trained on a specific problem format in one of three instructional conditions: speech alone, gesture alone, or speech + gesture (see *Training* for more detail on these conditions). Following training, participants completed a posttest consisting of problems identical in structure to trained problems, as well as several items designed to test participants' ability to generalize their knowledge to a new problem structure. The study thus has one independent variable with three levels: type of instruction (speech alone, gesture alone, speech + gesture) and one dependent variable: performance at posttest. We also explored reaction time as another dependent variable.

Rule Introduction: After completing informed consent, participants watched five videos that showed equations depicting the rules underlying the symbolic system while a voiceover explained each rule (e.g., "The order of two symbols does not change the result;" see Table 6 for a description of the rules of the system, and Appendix C for the exact scripts used in each video).

Table 6. The rules for combining symbols in the system from Kaminski et al. (2008) with updated symbols.

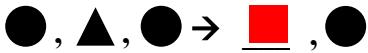
Rule 1: The order of two symbols does not change the result.
Rule 2: When a symbol combines with \blacksquare , the result will always be the other symbol.
Rule 3: $\bullet, \blacktriangle \rightarrow \blacksquare$
Rule 4: $\bullet, \bullet \rightarrow \blacktriangle$
Rule 5: $\blacktriangle, \blacktriangle \rightarrow \bullet$
Rule 6: The result of more than two symbols does not depend on which two symbols combine first.

Pretest: Following the rule introduction videos, participants completed a pretest designed to assess their baseline understanding of the rules and their ability to apply the rules to novel problems. The pretest consisted of 21 multiple-choice problems in which participants had to select one of three possible symbols to correctly solve an equation. These equations included simple 2-symbol combinations designed to test participants' knowledge of the rules ("rule-check" problems), 8 "typical format" problems (problems similar in structure to traditional math equivalence problems used in gesture studies with children, containing 3 addends on the left side of the equation and a blank and an equal added on the right side of the equation) and 8 generalization problems (problems that could not be solved using a simple grouping strategy; see Table 7 for examples of each problem type). Problems were presented in a fixed order for all participants: rule-check followed by typical format followed by generalization problems.

Training: Participants watched 8 videos of a teacher solving typical format problems. In all videos, the teacher explained how to solve the problem either in speech alone, with gesture accompanied by vacuous speech, or in speech with an accompanying gesture. In the two conditions with meaningful speech, the teacher used an equalizer strategy in speech: “I want to make one side equivalent to the other side;” in the gesture alone condition, the teacher said a neutral phrase: “Take a look at this problem.” In the two conditions containing gesture, the teacher performed an equalizer strategy in gesture (left-hand underlining the left side of the equation from left to right, then right to left, followed by a pause, followed by the right-hand underlining the right side of the equation from right to left, then left to right). Following each video, participants solved a new problem of the same structure shown in the video, and were given feedback on their response.

Posttest: After training, participants completed a posttest identical in structure to the pretest. The rule-check problems on the posttest were the same as those on the pretest (as there are a limited number of problems of this structure), but the order of these problems was randomized. The 8 typical format problems and 8 generalization problems on the posttest differed from those shown on the pretest. As on the pretest, problems were presented in a fixed order for all participants: rule-check followed by typical format followed by generalization problems.

Table 7. Examples of problem types (solutions appear in red).

Rule-learning: two-symbol combinations (5 items)	
Trained format: contain equal addends on each side of the equation (8 items)	
Generalization: require two operations on the left side of the equation in order to solve (8 items)	

Procedure: The experiment was conducted in an online session, lasting an average of 31 minutes. Participants were given informed consent prior to beginning the study, and they completed the entire study on their home computer. During the pretest, training, and posttest phases of the study, participants pressed the “J,” “K”, or “L” keys, which corresponded to the answer choices for each item, to make their responses. Participants’ reaction time for each trial for each phase of the study was also recorded. Participants were paid \$2.50 for completing the study.

Results

Exclusion Criteria: In exploring differences in reaction time between conditions and controlling for reaction time in evaluating performance, we wanted to exclude participants with reaction times far outside the average reaction time throughout the study. We first looked at trials on which participants had extremely high response times compared to the mean response time, as such response times likely indicated that participants took a short break from the study or got distracted on a particular trial (given that participants completed the study on their home computers rather than in a lab setting, this is a reasonable assumption). The median reaction time for each item throughout the study was 5.5 seconds and the mean reaction time was 12.7 seconds

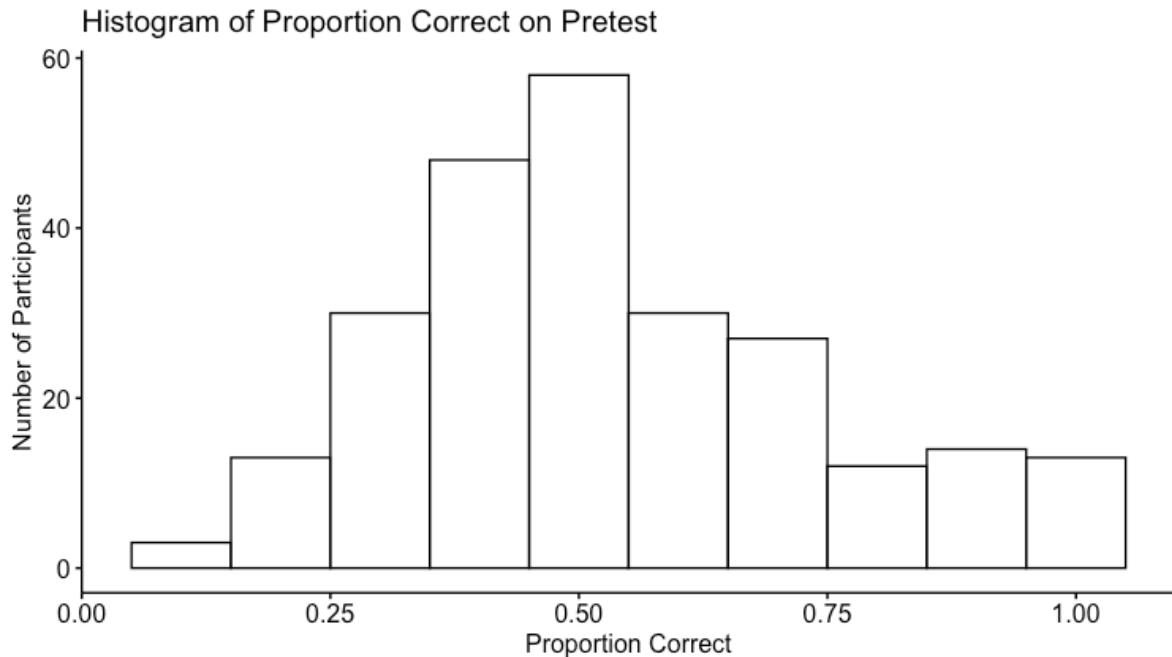
($SD = 131.6$ seconds). We excluded 61 participants who had a reaction time greater than one standard deviation from the mean reaction time on one or more trials.

For this new sample, we then looked at reaction time for pretest trials, and excluded data for trials on which participants had a reaction time more than 3 standard deviations above the mean reaction time for that phase of the study, as well as trials for which reaction time was recorded as “0” seconds by Qualtrics, likely indicating that participants’ response time for these trials was below the sensitivity threshold for Qualtrics’ timing questions. The mean reaction time for pretest trials was 8.0 seconds ($SD = 7.5$ seconds), and thus, we excluded trials on which participants took more than 30.6 seconds to respond. This resulted in excluding 2.1% of pretest trials (110 trials).

After these reaction time exclusions, we looked at pretest accuracy across all problem types, using proportion correct our accuracy measure. Mean proportion correct was 0.54 ($SD = 0.20$), and accuracy was somewhat normally distributed around this mean (see Figure 11). Given that participants who performed very well on the pretest had little to gain from the training videos, we excluded participants who performed better than one standard deviation above the mean – thus, we excluded an additional 41 participants whose proportion of problems solved correctly at pretest was greater than 0.74.

Based on this exclusion criteria, our final sample included 207 participants (speech alone: $n = 74$, gesture alone: $n = 65$, speech+gesture: $n = 68$).

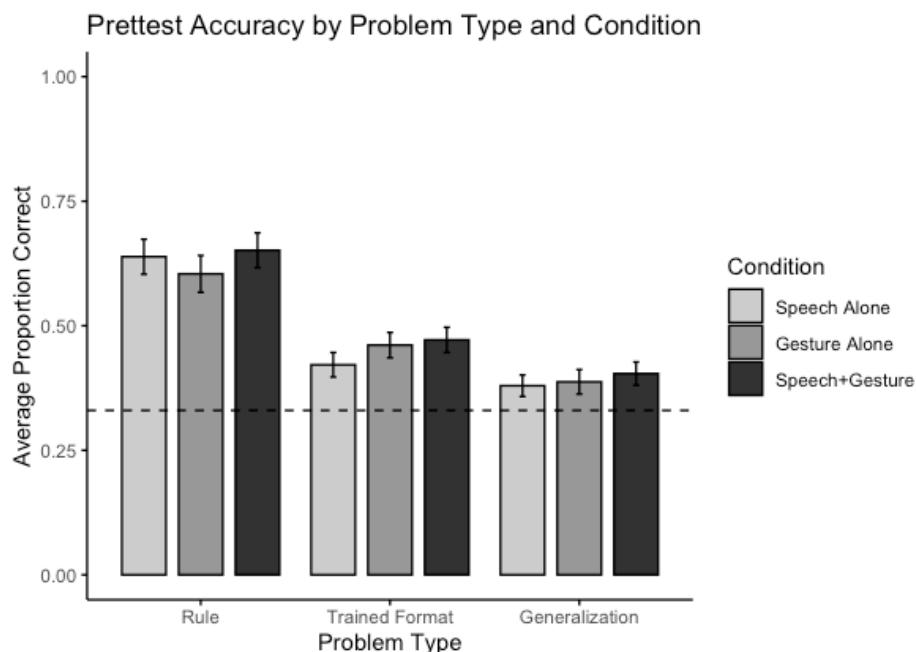
Figure 11. Histogram of proportion correct at pretest.



Pretest Accuracy: Having established a final sample, we first wanted to confirm that there were no significant differences on pretest accuracy by randomly-assigned condition. Because participants had not yet received any training instruction other than the rule videos (which were the same for all participants), we did not expect to see differences in pretest accuracy or reaction time on problem-solving by condition. To test this hypothesis, we ran a mixed-effects binomial logistic regression model with condition (speech alone, gesture alone, or speech+gesture), problem type (rule, trained format, or generalization), and an interaction as main effects; reaction time as a covariate; participant as a random effect; and proportion correct at pretest as the outcome variable. An analysis of variance of the model revealed no main effect of condition ($\chi^2 (2) = 2.1, p = 0.3$), a main effect of problem type ($\chi^2 (2) = 144.3, p < 0.000$), and no interaction between condition and problem type ($\chi^2 (4) = 3.0, p = 0.6$). Participants performed better on rule problems compared to trained format problems ($\beta = 0.6, z = 3.9, p < 0.000$) and generalization problems ($\beta = 0.9, z = 5.9, p < 0.000$), and better on trained problems compared to

generalization problems ($\beta = 0.3$, $z = 2.4$, $p = 0.02$). There was not a significant association between reaction time and pretest accuracy ($\beta = 0.002$, $z = 0.2$, $p = 0.9$). Thus, the analysis gave us confidence that random-assignment had been successful in ensuring that participants did not differ by training condition in their ability to solve the pretest problems prior to beginning the training phase of the study.

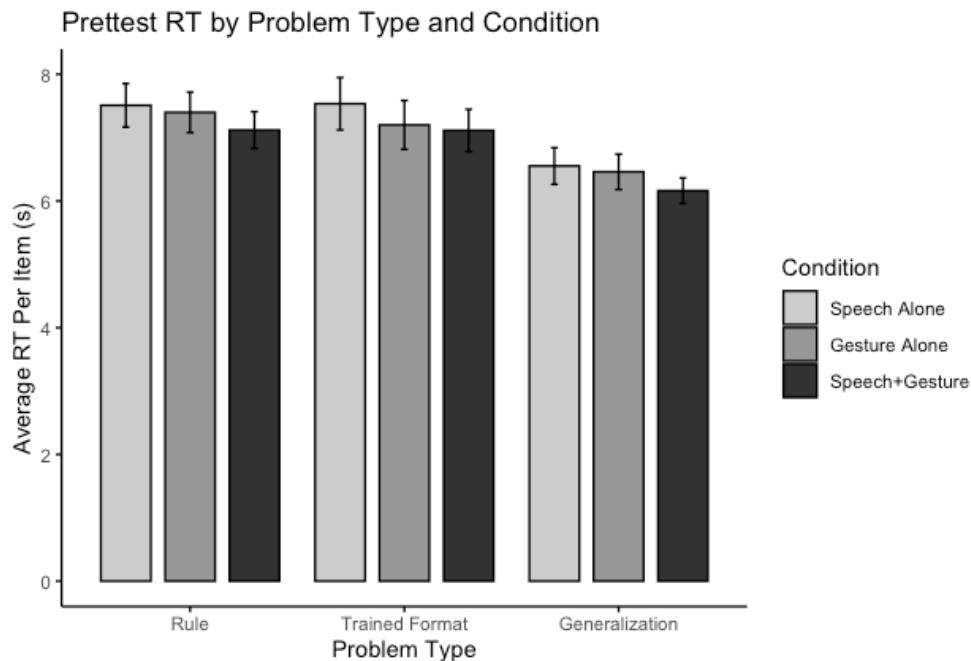
Figure 12. Pretest accuracy by problem type and condition.



Pretest Reaction Time: To explore whether participants responded more quickly on certain problem types and whether reaction time varied by condition, we ran a generalized linear model with condition (speech alone, gesture alone, or speech+gesture), problem type (rule, trained format, or generalization), and an interaction as main effects to predict mean reaction time. An analysis of variance of the model revealed no effect of condition ($\chi^2 (2) = 2.3$, $p = 0.3$), a main effect of problem type ($\chi^2 (2) = 16.2$, $p < 0.00$), and no interaction between condition and problem type ($\chi^2 (4) = 0.2$, $p = 1.0$). Participants performed significantly faster on generalization problems than on rule problems ($\beta = -0.9$, $t = -2.0$, $p < 0.05$). This analysis gave us confidence

that participants did not differ by training condition in their problem-solving speed prior to beginning the training phase of the study.

Figure 13. Pretest reaction time by problem type and condition.

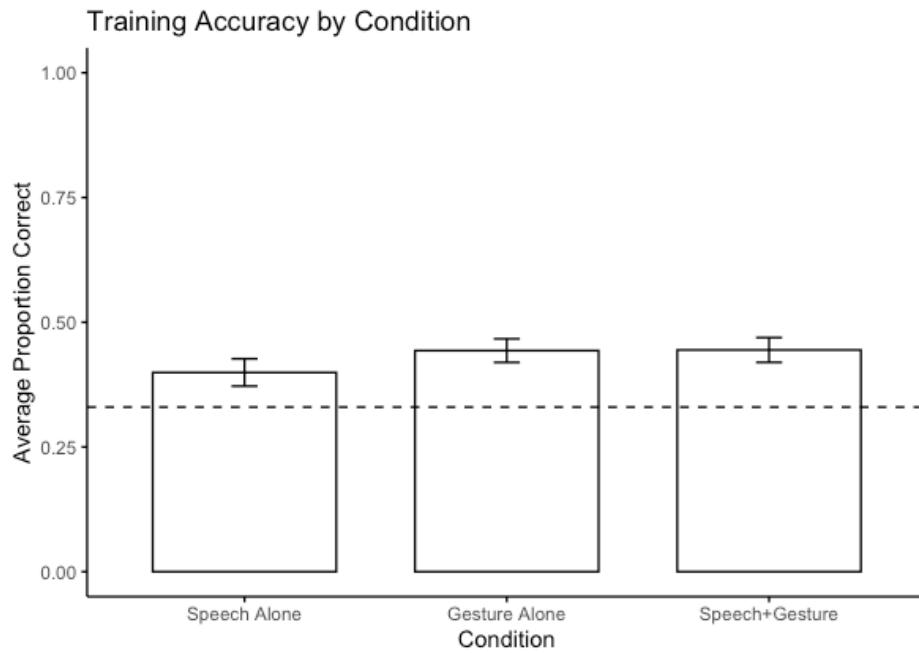


Training Accuracy: Prior to looking at accuracy on training, we first looked at reaction time for training trials, and excluded data for trials on which participants had a reaction time more than 3 standard deviations above the mean reaction time for this phase of the study, as well as trials for which reaction time was recorded as “0” seconds by Qualtrics, likely indicating that participants’ response time for these trials was below the sensitivity threshold for Qualtrics’ timing questions. The mean reaction time for training trials was 8.6 seconds ($SD = 11.0$ seconds), and thus, we excluded trials on which participants took more than 41.4 seconds to respond. This resulted in excluding 2.3% of training trials (38 trials).

To explore training accuracy by condition, we ran a mixed-effects binomial logistic regression model with condition (speech alone, gesture alone, or speech+gesture) as a main effect; reaction time as a covariate; participant as a random effect; and proportion correct on

training items as the outcome variable. An analysis of variance of the model found no main effect of condition ($\chi^2 (2) = 2.8, p = 0.2$). Reaction time was significantly positively associated with training accuracy ($\beta = 0.03, z = 2.5, p = 0.01$).

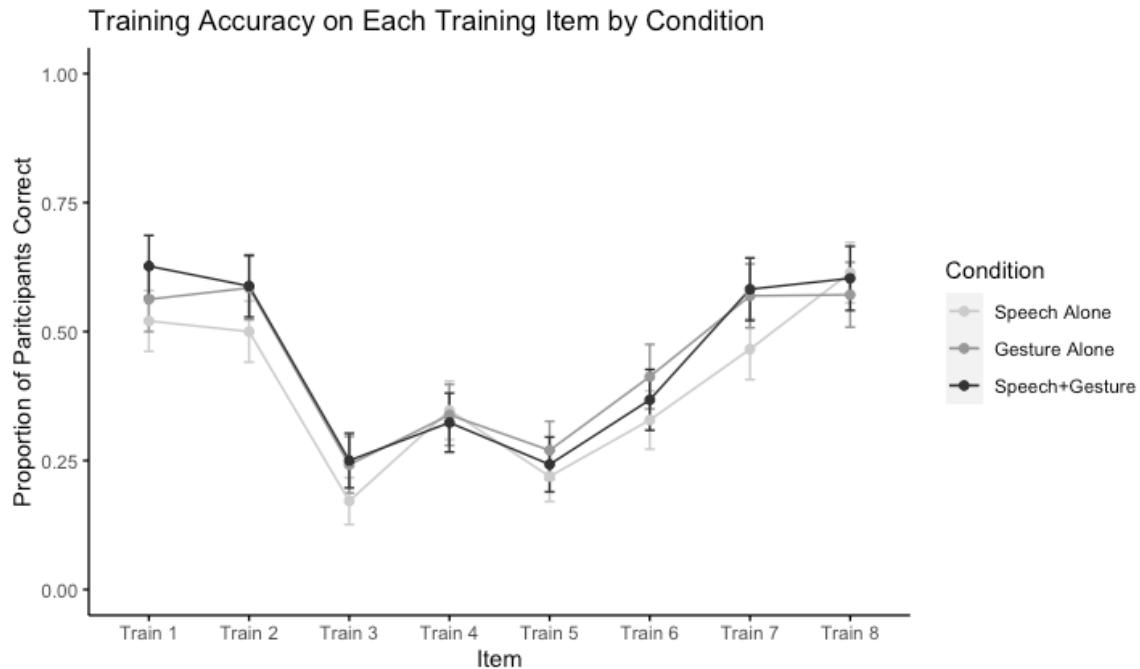
Figure 14. Training accuracy by condition.



We also explored whether participants improved on training items across the 8 problems they solved during training, and whether this varied by condition. To test this, we ran another mixed-effects binomial logistic regression model with condition (speech alone, gesture alone, or speech+gesture), training item (items 1-8), and an interaction as main effects; reaction time as a covariate; participant as a random effect; and accuracy on each item (0, 1) as the outcome variable. An analysis of variance of the model found no effect of condition ($\chi^2 (2) = 2.7, p = 0.3$), a significant effect of item ($\chi^2 (7) = 131.0, p < 0.000$), and no interaction ($\chi^2 (14) = 5.2, p = 1.0$). Reaction time was not significantly associated with training accuracy across items ($\beta = 0.03, z = 2.5, p = 0.01$). Across conditions, participants showed a u-shaped pattern of performance, performing well on the first two training items, struggling with items 3-6, and returning to high

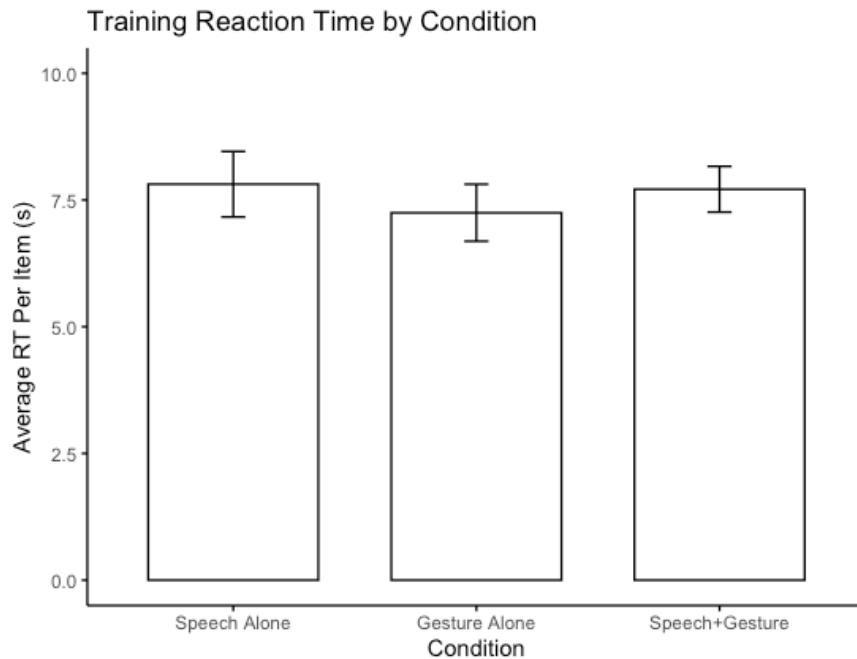
levels of performance on items 7 and 8. Items were equated for difficulty, so it is unclear why participants struggled with particular training items compared to others.

Figure 15. Training performance on each training trial by condition.



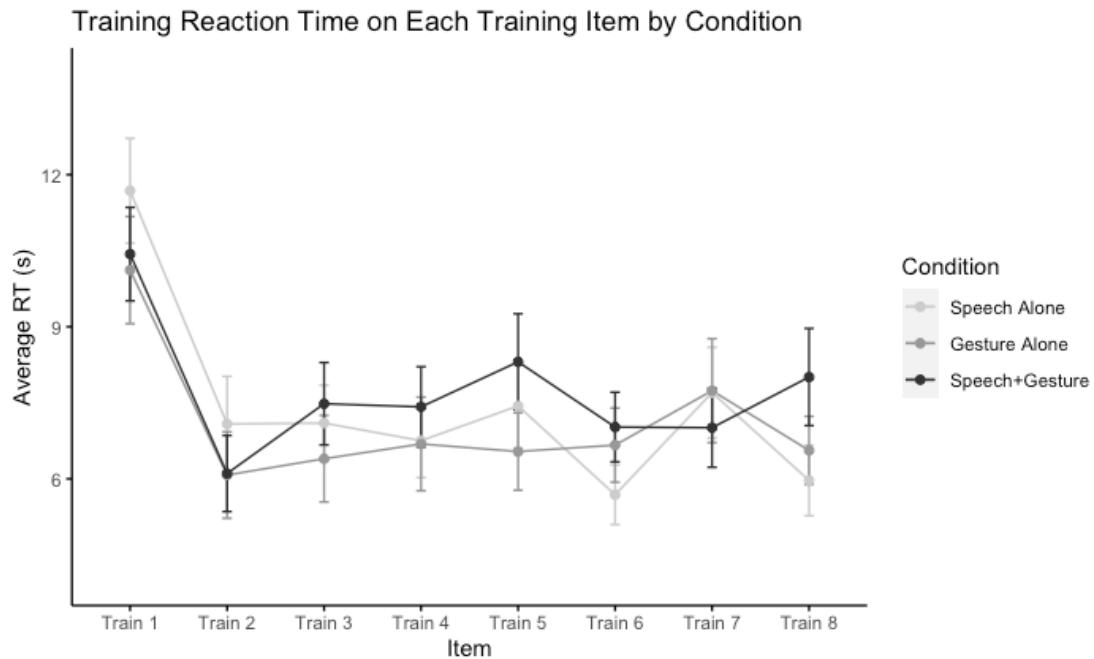
Training Reaction Time: To explore whether reaction time on training problems varied by condition, we ran a generalized linear model with condition (speech alone, gesture alone, or speech+gesture) as a main effect to predict reaction time. An analysis of variance of the model revealed no main effect of condition ($\chi^2 (2) = 0.6, p = 0.8$).

Figure 16. Training reaction time by condition.



We also explored whether participants improved on reaction time across the 8 problems they solved during training, and whether this varied by condition. To test this, we ran another generalized linear model with condition (speech alone, gesture alone, or speech+gesture), training item (items 1-8), and an interaction as main effects to predict reaction time. An analysis of variance of the model revealed no effect of condition ($\chi^2 (2) = 2.1, p = 0.3$), a significant effect of item ($\chi^2 (7) = 59.5, p < 0.000$), and no interaction ($\chi^2 (14) = 9.2, p = 0.8$). Participants were slower on item 1 compared to all other items, possibly as a result of acclimating to the training procedure of first watching a video, followed by problem-solving.

Figure 17. Training reaction time on each training trial by condition.

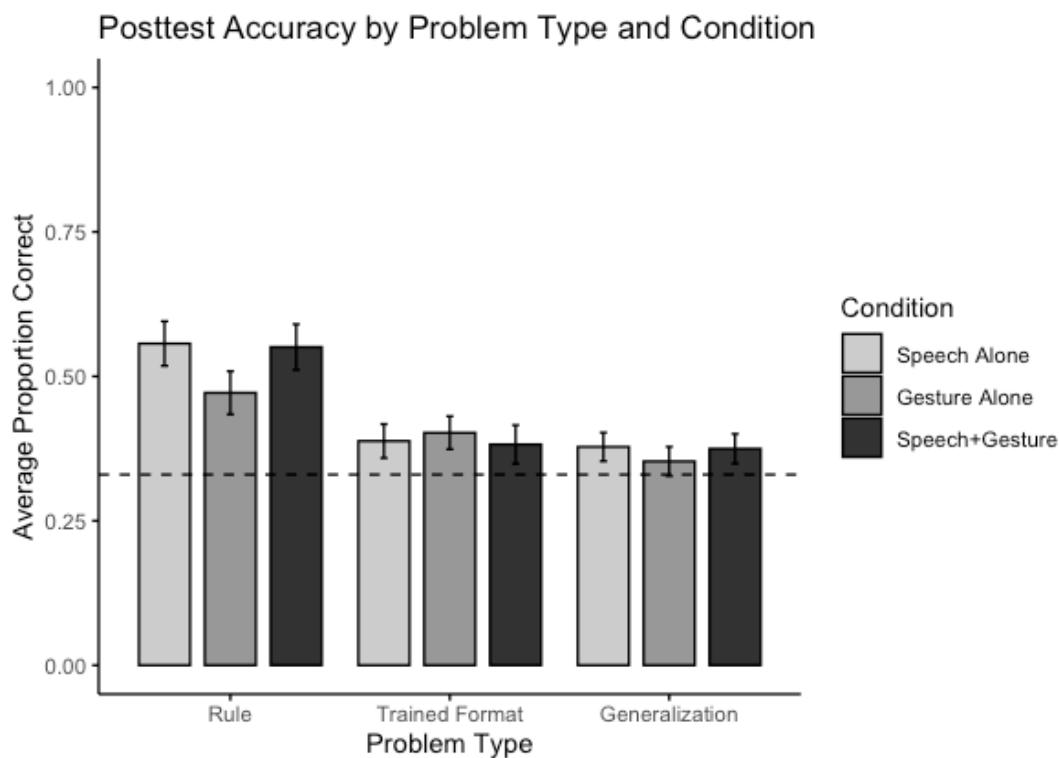


Posttest Accuracy: Prior to looking at accuracy at posttest, we first looked at reaction time for posttest items, and excluded data for trials on which participants had a reaction time more than 3 standard deviations above the mean reaction time for this phase of the study, as well as trials for which reaction time was recorded as “0” seconds by Qualtrics, likely indicating that participants’ response time for these trials was below the sensitivity threshold for Qualtrics’ timing questions. The mean reaction time for posttest items was 6.8 seconds ($SD = 5.7$ seconds), and thus, we excluded trials on which participants took more than 23.9 seconds to respond. This resulted in excluding 1.8% of posttest trials (78 trials).

To explore whether participants learned differently depending on the instruction they received during training, we ran a mixed-effects binomial logistic regression model with condition (speech alone, gesture alone, or speech+gesture), problem type (rule, trained format, or generalization), and an interaction as main effects; reaction time as a covariate; participant as a random effect; and proportion correct at posttest as the outcome variable. An analysis of variance

of the model revealed no effect of condition ($\chi^2 (2) = 0.6, p = 0.7$), a significant effect of problem type ($\chi^2 (2) = 77.5, p < 0.000$), and no interaction ($\chi^2 (4) = 6.5, p = 0.2$). Reaction time was not significantly associated with posttest accuracy ($\beta = -0.002, z = -0.1, p = 0.9$). Participants performed significantly better on rule problems compared to trained format problems ($\beta = 0.8, z = 5.4, p < 0.000$) and generalization problems ($\beta = 0.8, z = 5.7, p < 0.000$).

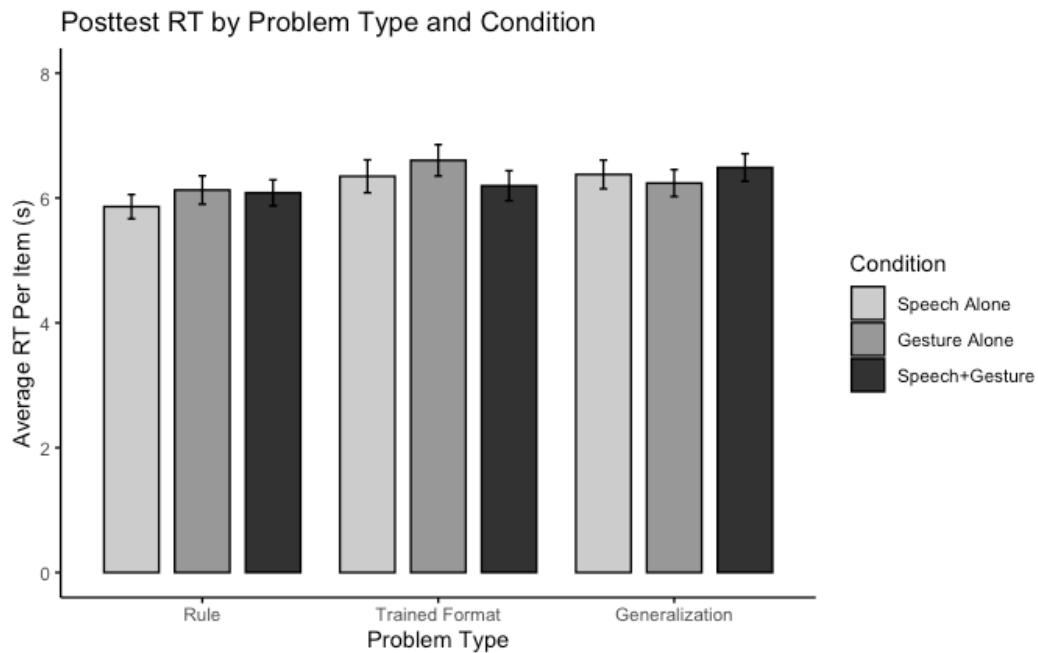
Figure 18. Posttest accuracy by problem type and condition.



Posttest Reaction Time: To explore whether participants responded more quickly on certain problem types at posttest and whether this varied by condition, we ran a generalized linear model with condition (speech alone, gesture alone, or speech+gesture), problem type (rule, trained format, or generalization), and an interaction as main effects to predict reaction time on each trial. An analysis of variance of the model revealed no effect of condition ($\chi^2 (2) = 0.5, p = 0.8$), a trending effect of problem type ($\chi^2 (2) = 4.9, p = 0.09$), and no interaction ($\chi^2 (4) = 2.4, p$

= 0.7). The trending effect of problem type did not emerge in the model, but beta values suggested that participants were marginally faster on rule problems compared to trained format problems ($\beta = -0.4$, $t = -1.9$, $p = 0.05$) and generalization problems ($\beta = -0.3$, $t = -1.9$, $p = 0.06$).

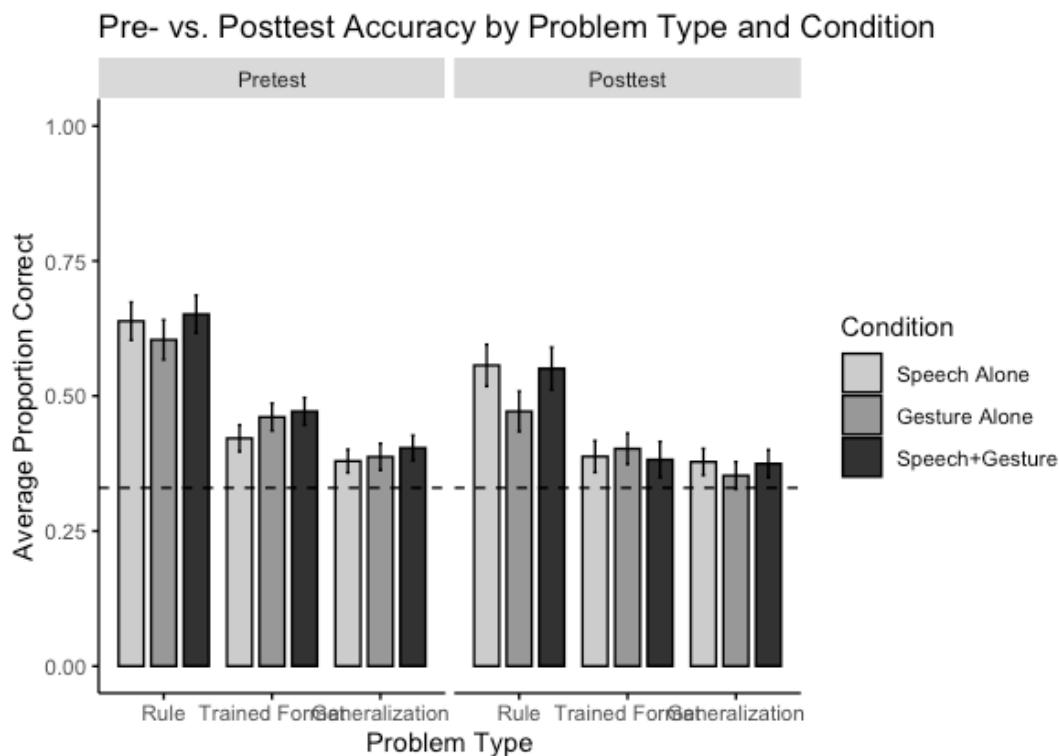
Figure 19. Posttest reaction time by problem type and condition.



Comparing Pre- to Posttest Accuracy: To explore whether participants improved from pre- to posttest depending on the instruction they received during training, we ran a mixed-effects binomial logistic regression model with condition (speech alone, gesture alone, or speech+gesture), problem type (rule, trained format, or generalization), assessment (pretest or posttest), and interactions as main effects; reaction time as a covariate; participant as a random effect; and proportion correct at each assessment as the outcome variable. An analysis of variance of the model revealed no effect of condition ($\chi^2 (2) = 0.7$, $p = 0.7$), a significant effect of problem type ($\chi^2 (2) = 217.2$, $p < 0.000$), a significant effect of assessment ($\chi^2 (1) = 28.9$, $p < 0.000$), a trending interaction of condition and problem type ($\chi^2 (2) = 8.3$, $p = 0.08$), and a

significant interaction of problem type and trial type ($\chi^2 (2) = 9.5, p = 0.009$). Reaction time was not significantly associated with accuracy across the pre- and posttests ($\beta = -0.01, z = -1.0, p = 0.3$). Participants across conditions performed significantly worse at posttest compared to pretest on rule problems ($\beta = -0.4, z = -3.1, p = 0.002$) and marginally worse on trained format problems ($\beta = -0.2, z = -1.7, p = 0.09$). Participants performed significantly above chance on all problem types at pretest and posttest (pretest: rule problems ($t(206) = 14.7, p < .000$), trained format problems ($t(206) = 8.3, p < .000$), generalization problems ($t(206) = 4.5, p < .000$); posttest: rule problems ($t(206) = 8.9, p < .000$), trained format problems ($t(206) = 3.5, p < .001$), generalization problems ($t(206) = 2.7, p < .01$).

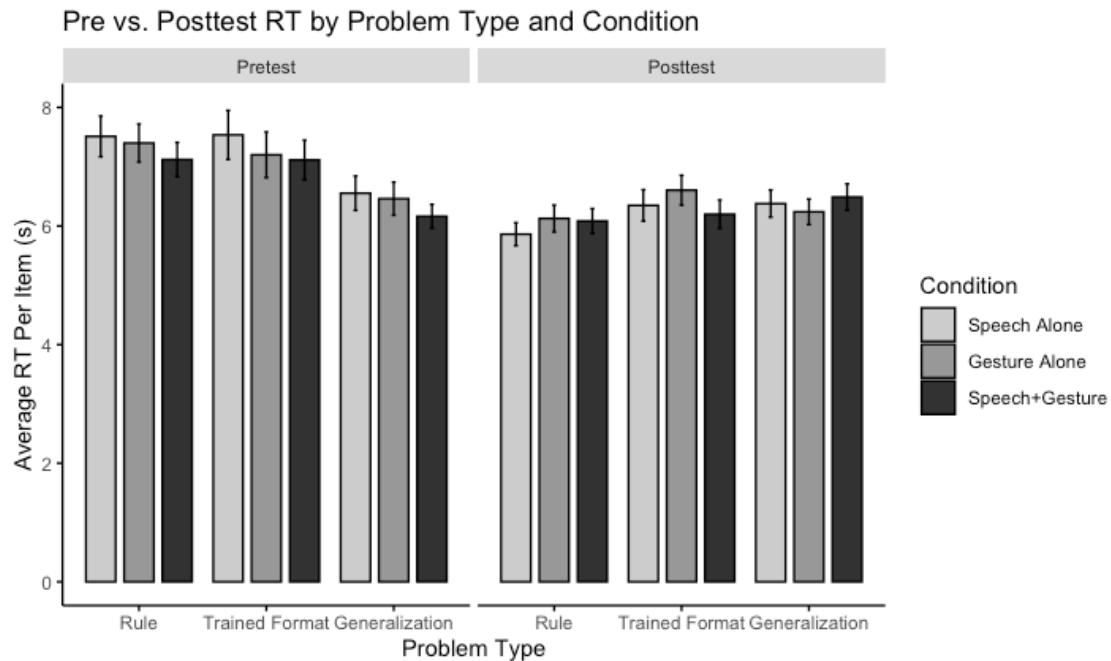
Figure 20. Pretest vs. posttest accuracy by problem type and condition.



Comparing Pre- to Posttest Reaction Time: The decrease in accuracy between pre- and posttest across all training conditions on rule and trained format problems is very surprising, as it

suggests that the instruction participants received during training not only failed to help them improve their understanding, but may have actually hindered their ability to acquire knowledge about the symbolic system. As a different measure of performance, we examined whether participants were able to solve problems at posttest more quickly compared to pretest. We ran a generalized linear model with condition (speech alone, gesture alone, or speech+gesture), problem type (rule, trained format, or generalization), assessment (pretest or posttest), and interactions as main effects and reaction time on each assessment item as the outcome variable. An analysis of variance of the model revealed no effect of condition ($\chi^2 (2) = 1.2, p = 0.5$), a significant effect of problem type ($\chi^2 (2) = 8.0, p = 0.02$), a significant effect of trial type ($\chi^2 (1) = 32.6, p < 0.000$), and an interaction of problem type and trial type ($\chi^2 (2) = 16.9, p < 0.00$). Participants were significantly faster at posttest vs. pretest on rule problems ($\beta = -1.3, t = -4.0, p < 0.000$) and trained format problems ($\beta = -0.9, t = -2.7 p = 0.01$). This faster reaction time from pre- to posttest on rule and trained format problems may partially explain participants' worse performance at posttest compared to pretest on these problem types. Even if participants gained some useful information from the training phase of the study, they may have failed to apply this knowledge if they were answering the posttest items too quickly to apply any learned problem-solving strategies.

Figure 21. Pretest vs. posttest reaction time by problem type and condition.



Discussion

The results of the present study are inconclusive. We did not find significant differences in problem-solving accuracy or speed by training condition on either the training or posttest phases of the study. From pre-to-posttest, participants' accuracy on rule and trained format problems actually worsened. This result is very surprising, particularly for the speech+gesture condition, based on a large body of research showing that instruction with speech and gesture promotes better learning than instruction with speech alone on a variety of concepts (e.g., Church, Ayman-Nolley, & Mahootian, 2004; Singer & Goldin-Meadow, 2005; Cook & Goldin-Meadow, 2006; Ping & Goldin-Meadow, 2008; Valenzano, Alibali, & Klatzky, 2003). Participants performed above chance on all problem types at pretest and posttest, indicating that they did learn from the introductory rule videos; however, their worse performance at posttest compared to pretest indicates that participants forgot their knowledge from these videos over the

course of the study and/or that the training videos were actually detrimental to their understanding.

Varied performance at pretest suggests that participants gleaned different levels of knowledge about the symbolic system from the introductory rule videos, which may have affected how participants across the different conditions benefited from the training phase of the study. It is possible that having a weak grasp of the rules at the beginning of the study may have prevented some participants from gaining meaningful information from subsequent training videos. One important amendment to a future study design could be to make participants reach a certain “rule understanding” criterion prior to training, perhaps by having participants repeatedly watch the rule videos and test their understanding of the rules via two-symbol combination problems until they demonstrate above chance performance on these problems.

Furthermore, solving trained format and generalization problems at pretest gave participants an opportunity prior to training to practice applying their knowledge from the rule videos. Though participants did not receive feedback during the pretest, this experience with problem-solving may have given participants insight into the different problem structures and experience with applying any knowledge they had gained about the rules from the introductory videos. Another possibility is that participants found the pretest cognitively draining. In either case, participants may have benefited less from training than if they had not completed the pretest. The intent behind including a pretest in our design was to provide a measure of participants’ baseline knowledge prior to training so that posttest performance could serve as an informative measure of how much participants benefited from training. The fact that participants performed worse at posttest on rule and trained format problems while simultaneously answering these problems more quickly compared to at pretest suggests that training across the different

conditions may have fatigued participants. Perhaps giving participants the opportunity to take a short break or providing them with a reminder of the rule structure underlying the symbol system after training may have led to improvement at posttest, and measurable differences between training conditions.

It would be interesting to replicate our study design using a different gesture, such as a grouping gesture similar to that used in math equivalence learning studies with children (e.g., Goldin-Meadow et al., 2009; Novack et al., 2014). Such a gesture would specifically highlight the rule structure underlying all the problems shown to participants by demonstrating how two symbols can be combined to form another symbol. In contrast to the equivalence strategy used in the present study, which depicted a general principle for solving all problem types, a grouping gesture may provide learners with a more salient, but still nondeclarative representation of an algorithmic problem-solving strategy. Our hypothesis was that gesture may promote learning in this context by conveying nondeclarative knowledge – however, the *type* of knowledge conveyed by gesture may be particularly important for it to positively impact learning. We explore this question in more detail in Chapter 4 of this dissertation.

Finally, we were unable to collect a large amount of followup data from participants in the present study. We provided participants with the option to complete a followup posttest 24-hours after the immediate posttest, but only 48 participants completed this followup, and of these 48, only 28 were included in our final sample for day 1 of the study, resulting in a followup sample that is far too small to glean meaningful findings from. Amazon MTurk does not provide a researcher-friendly system for longitudinal studies (e.g., it is against MTurk policy to collect participant email addresses to send them a followup survey), and thus, attrition for the followup study was high. Given that the benefits of observing gesture on learning have been shown to

emerge and persist over time in multiple previous studies (e.g., Cook, Duffy, & Fenn; 2013; Wakefield et al., 2019; Congdon et al., 2017) it is important that future research prioritizes collecting useful followup data in order to fully understand gesture's impact on learning for different concepts.

Of the 28 participants who completed the followup, 14 saw gesture during training. The participants completed a questionnaire at the end of the followup and were asked if they noticed gestures in the videos they watched the previous day. 7 of the participants who saw gesture indicated that they noticed the gestures. Though this is a very small representation of our sample, this result indicates that many participants may have failed to pay attention to the gestures during training or did not find them to be informative. We know from prior work that gesture's impact on learning can be unconscious (e.g., Ellis Weismer & Hesketh, 1993), but to our knowledge, this is one of the first studies in which participants were asked to view and learn from videos containing gestures outside of a laboratory setting. Previous work has found that children and adults can learn from videos of teachers gesturing (e.g., Church, Ayman-Nolley, & Mahootian, 2004; Wakefield et al., 2018b; Cook & Fenn, in prep.), but these studies have been conducted in the presence of an experimenter, which may have motivated learners to pay attention to the videos, and by extension, the gestures conveyed in them. In contrast, the participants in our study watched the videos on their home computers, and thus, may have been less engaged with the study than they would have been in a lab setting, or have encountered distractions that are unlikely to occur during a controlled experiment. Participants in Study 1 were able to benefit from gesture instruction under similar circumstances as the participants in Study 2, however, Study 1 involved much more training and repeated exposure to gesture than did Study 2. As the world rapidly transitions to online educational formats (particularly in the midst of the Covid-19

crisis), it will be important for researchers and educators to explore how to improve online instruction, and how to translate beneficial learning devices like gesture to online platforms.

CHAPTER FOUR: VARYING GESTURAL KNOWLEDGE AND SPEECH

Introduction

Learners benefit from observing teachers' gestures as they explain a concept. Instruction with gesture can facilitate understanding and retention of a variety of concepts for both children and adult learners (e.g., Singer & Goldin-Meadow, 2005; Ping & Goldin-Meadow, 2008; Kelly, McDevitt, & Esch, 2009; Cook, Duffy, & Fenn, 2013; Congdon et al., 2017; Wakefield et al., 2019; Cook & Fenn, in prep.). Gesture likely benefits learning through many different mechanisms, such as by activating a learner's own motor system (Ping et al., 2014), and by complimenting information conveyed in speech (Singer & Goldin-Meadow, 2005; Wakefield et al., 2018).

In the present study, we explore this third mechanism in the context of implicit and explicit learning processes. Specifically, we argue that gesture may facilitate learning by conveying unspoken, nondeclarative knowledge at the same time as spoken, declarative knowledge, thus helping learners transition from a state of implicit to explicit knowledge of a concept. Some work suggests that gestural information may be acquired implicitly, as gesture can change a learner's knowledge of a concept even if the learner is unable to verbally report that the knowledge came from gesture (Broaders & Goldin-Meadow, 2010; Gurney, Pine & Wiseman, 2013; Kelly et al., 1999; Thompson, 1995; Kelly et al., 2004). At the same time, gesture has a privileged relationship with speech, the gold standard for declarative knowledge. Unlike other types of movements, gesture forms a coherent system of communication with the speech it accompanies (Bernardis & Gentilucci, 2006; Church, Kelly, Holcombe, 2014; Loehr, 2007; McNeill, 1992; Nobe, 2000). Importantly, information conveyed in a speaker's gestures

can either reiterate (match) or differ from (mismatch) the information conveyed in speech (Goldin-Meadow, 2003; Goldin-Meadow, Alibali, & Church, 1993).

Intentionally produced gesture-speech mismatches have been shown to promote learning in some contexts. For example, Singer and Goldin-Meadow (2005) taught children to solve math equivalence problems (e.g., $5 + 4 + 2 = \underline{\quad} + 2$) by observing an experimenter perform different speech and gesture strategies during problem-solving explanations. In all learning conditions, children saw and heard the experimenter give an equalizer strategy in speech: “I want to make one side equal to the other side.” In a *speech alone* condition, children saw and heard this equalizer strategy in speech without gesture. In a *speech + matching gesture* condition, children saw and heard the experimenter perform a gesture that conveyed the same equalizer strategy as her spoken strategy (e.g., a sweeping gesture with the left hand first under the left side of the equation and then under the right with the right hand). Children in a *speech + mismatching gesture* condition saw the experimenter perform a gesture that conveyed a different strategy from the equalizer strategy conveyed in speech (e.g., points to the three numbers on the left side of the equation followed by a “take away” gesture under the number on the right side, an “add-subtract” strategy). In a final control group, children saw and heard the experimenter perform both the equalizer and the add-subtract strategies in speech (*speech alone – two strategies*). The study used a pretest-training-posttest design, and children in the speech + mismatching gesture condition showed significantly improved understanding of math equivalence compared to children in all other conditions, whereas children in the speech alone – two strategies condition performed significantly worse at posttest compared to children in all other conditions. Thus, children in the speech + mismatching gesture condition did not simply perform better because they received two strategies instead of one, but rather because one of these strategies was

conveyed uniquely in gesture. This finding suggests that unspoken knowledge conveyed by a teacher's gesture may make the teacher's spoken knowledge more salient to a learner, even when gesture conveys information that is different from speech. In this way, gesture may serve as a vehicle for conveying non-declarative knowledge.

The finding by Singer and Goldin-Meadow (2005) that learning through seeing gesture-speech mismatches promotes better learning of math equivalence than speech instruction alone has been replicated in subsequent studies. For example, Congdon and colleagues (2017) found that instruction with a simultaneous equalizer strategy in speech and an add-subtract strategy in gesture led to more effective and lasting learning than instruction with an equalizer strategy in speech alone. Wakefield and colleagues (2018b) similarly found that instruction with a simultaneous equalizer strategy in speech and a grouping strategy in gesture (e.g., for the problem $5 + 6 + 3 = \underline{\quad} + 3$, the teacher pointed to the 5 and 6 on the left side of the equation using a v-handshape, followed by a point to the blank on the right side) led to better learning than speech alone instruction. However, there has been little investigation into precisely *why* these types of mismatches benefit learning.

There are many possible mechanisms through which gesture-speech mismatches might promote learning. Mismatching gestures provide a second problem-solving strategy to an instructional context, and earning multiple problem-solving strategies is positively associated with cognitive change (Siegler, 1994). Additionally, work in social psychology has shown that learners who have discrepancies between their implicit and explicit attitudes or beliefs are motivated to consider new relevant information in order to resolve these discrepancies (Brinol, Petty & Wheeler, 2006; Rydell, McConnell, & Mackie, 2008). Based on this account, seeing gestures that convey information different from information conveyed in speech could create a

type of cognitive dissonance that encourages learners to engage in increased processing of information relevant to the task to be learned. Indeed, Hostetter (2011) conducted a meta-analysis of studies in which listeners' understanding of a message was compared when the message was presented with speech alone vs. speech+gesture and found that gesture's beneficial effect on comprehension was larger when gestures depicted information that was not completely redundant with speech.

One way to explore *how* mismatching benefits learning is by examining how the specific *type* of information conveyed by gesture in a gesture-speech mismatch impacts learning. Recall the add-subtract and equalizer strategies used by Singer and Goldin-Meadow (2005). "Add-subtract" offers an algorithm for solving problems, whereas "equalizer" explains a general principle underlying all correct solutions. Perry (1991) found that children benefit more from instruction on mathematical equivalence problems when taught the equalizer strategy as opposed to the add-subtract strategy, but perform worse if they are taught both the principle and algorithm strategies than if taught the principle strategy alone. At first glance, this finding seems to contradict previously observed benefits of learning multiple strategies. Singer and Goldin-Meadow (2005), however, showed that children could benefit from presentation of both strategies when one strategy was presented in gesture. Singer and Goldin-Meadow (2005) proposed that because gesture is less explicit than speech, the information it conveys may be less intrusive to a learner than information conveyed in speech. Specifically, gesture may help children understand the relationship between principle and algorithm by presenting the strategies simultaneously and not sequentially, as is required when both strategies are taught in speech.

It is unclear from Singer and Goldin-Meadow (2005) whether the *modality* of each strategy impacted learning – specifically, whether the algorithm strategy had to be presented in

gesture, rather than speech, to be effective. Perhaps presenting learners with two strategies is more effective than one strategy when one strategy is presented in gesture, regardless of the type of strategy conveyed in each modality. In the present study, we explore how the unspoken, non-declarative knowledge conveyed in gesture interacts with the declarative knowledge conveyed in speech to impact learning by manipulating the type of information conveyed in each modality. We present adult learners with simultaneous speech + gesture instruction of either one or two strategies for solving novel symbolic math equivalence problems by manipulating whether gesture matches (one strategy) or mismatches (two strategies) the speech strategy. Additionally, we vary the type of strategy presented in gesture (principle vs. algorithm) and explore how these different strategies impact learning in gesture-speech matching vs. mismatching contexts. We use adapted versions of the equalizer (principle) and the grouping (algorithm) strategies used in previous research on math equivalence learning in children to create 4 learning conditions: 2 matching conditions (matching – speech principle + gesture principle; matching – speech algorithm + gesture algorithm) and 2 mismatching conditions (mismatching – speech algorithm + gesture principle; mismatching – speech principle + gesture algorithm). Our use of the grouping strategy (rather than the add-subtract strategy) as an algorithmic strategy will also allow us to replicate with adult learners the benefit of presenting this strategy in a mismatching context observed in previous work with children (e.g., Wakefield et al., 2018b).

We expect gesture to facilitate learning regardless of whether it reiterates spoken knowledge or conveys new unspoken knowledge. Based on the finding by Singer and Goldin-Meadow (2005) that children learned to solve math equivalence problems more effectively through two strategies when one of the strategies conveyed an algorithm (add-subtract) uniquely in gesture—that is, when gesture conveyed new unspoken knowledge—we expect the

mismatching speech principle + gesture algorithm condition to outperform both matching conditions. If gesture simply benefits learning when it conveys non-declarative knowledge that is different from speech regardless of the type of knowledge it conveys, the mismatching speech algorithm + gesture principle condition should perform as well as the other mismatching condition. However, an interaction between the *type* of knowledge conveyed and whether gesture reiterates this knowledge or conveys new knowledge will indicate that gesture may be most useful as a learning tool when it complements the declarative knowledge of a teacher's speech in a specific way. Based on findings that gesture instruction can promote generalization of a learned concept (Cook, Duffy, & Fenn, 2013; Congdon et al., 2017; Wakefield et al., 2018a), gesture may be particularly effective at conveying a complex, but "general" principle that can be applied across all problem types. At the same time, gesture is non-declarative, and therefore already outside of a learner's consciousness, and thus, a gesture that strengthens the "structural salience" (Reber et al., 1980) of the task by depicting a simple algorithm may also benefit learning. The results of our study are nuanced, but make important strides toward distinguishing between these possibilities.

Method

352 adult participants (*Mean age* = 36.5 yrs, SD = 10.8 yrs; 55% male, 44% female, <1% preferring not to say, <1% other gender) were recruited through Amazon Mechanical Turk and completed all components of the study via an online Qualtrics survey. The sample was racially and ethnically diverse (ethnicity: 66% identifying as not Hispanic, 30% Hispanic, 3% preferring not to say, 1% another ethnicity; race: 75% White, 11% Black, 7% Native American, 6% Asian, 1% preferring not to say, <1% other race). Like Study 2, Study 3 followed a pretest-training-posttest design. The study procedure as well as the rule introduction, pretest, posttest, and

follow-up assessments were identical to those used in Study 2. Following the pretest, participants were trained on the same problems used in Study 2, but the type of training varied by condition. Training videos varied on whether the teacher produced an equalizer (principle) strategy or a grouping (algorithm) strategy in gesture, and on whether this gesture conveyed the same information (match) or different information (mismatch) from the teacher's speech. Study 3 thus had two independent variables, each with two levels: match status (matching or mismatching) and type of gesture instruction (principle or algorithm), resulting in four different training conditions: *matching – gesture principle*; *matching – gesture algorithm*, *mismatching – gesture principle*, *mismatching – gesture algorithm*.

Training: Participants watched 8 videos of a teacher solving typical format problems. In all 8 videos, the teacher explained how to solve the problem while producing simultaneous strategies in speech and gesture. In the *matching – speech principle + gesture principle* condition, the teacher produced an *equalizer* strategy in speech ("I want to make one side equivalent to the other side") while producing a matching *equalizer* strategy in gesture (left-hand underlining the left side of the equation from left to right, then right to left, followed by a pause, followed by the right-hand underlining the right side of the equation from right to left, then left to right). In the *matching – speech algorithm + gesture algorithm* condition, the teacher produced a grouping strategy in speech (this speech varied slightly depending on the symbols that could be combined on the left side of the equation to solve the problem, e.g., "On the left side, circle combined with circle makes triangle") while producing a matching *grouping* strategy in gesture (V-handshape under two adjacent symbols on the left side of the equation and point to blank on the right side of the equation). In the *mismatching – speech algorithm + gesture principle* condition, the teacher produced a *grouping* strategy in speech while producing an

equalizer strategy in gesture. In the *mismatching – speech principle + gesture algorithm* condition, the teacher produced an *equalizer* strategy in speech while producing a *grouping* strategy in gesture. In all conditions, the teacher concluded the speech strategy by stating the solution to the problem (e.g., “The answer is triangle”).

Results

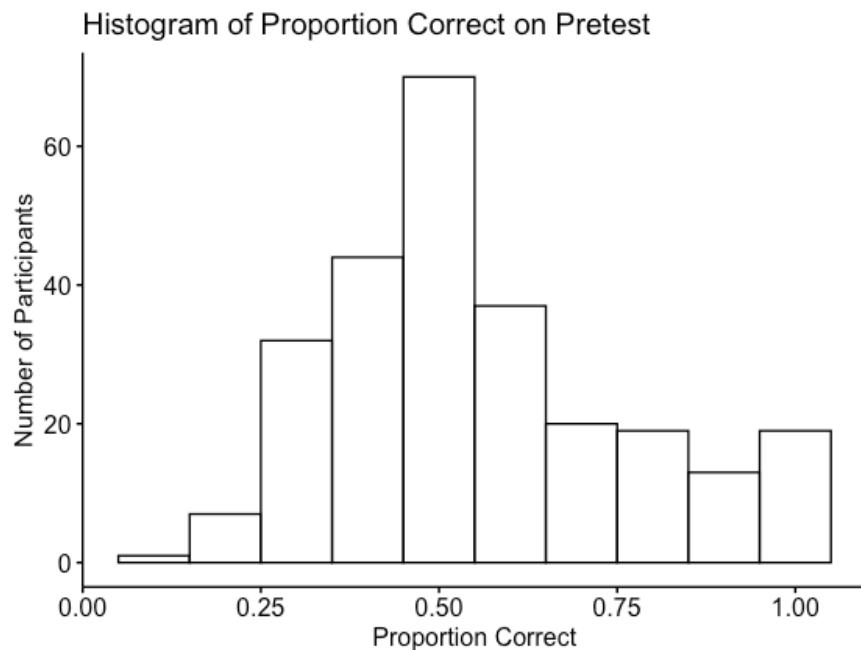
Exclusion Criteria: In exploring differences in reaction time between conditions and controlling for reaction time in evaluating accuracy, we wanted to exclude participants with reaction times far outside the average reaction time throughout the study. We first looked at trials on which participants had extremely high response times compared to the mean response time, as such response times likely indicated that participants took a short break from the study or got distracted on a particular trial (given that participants completed the study on their home computers rather than in a lab setting, this is a reasonable assumption). The median reaction time for each item throughout the study was 5.5 seconds and the mean reaction time was 10.8 seconds ($SD = 66.0$ seconds). We excluded 91 participants who had a reaction time greater than one standard deviation from the mean reaction time on one or more trials.

For this new sample, we then looked at reaction time for pretest trials, and excluded data for trials on which participants had a reaction time more than 3 standard deviations above the mean reaction time for that phase of the study, as well as trials for which reaction time was recorded as “0” seconds by Qualtrics, likely indicating that participants’ response time for these trials was below the sensitivity threshold for Qualtrics’ timing questions. The mean reaction time for pretest trials was 7.9 seconds ($SD = 6.4$ seconds), and thus, we excluded trials on which participants took more than 27.2 seconds to respond. This resulted in excluding 2.0% of pretest trials (108 trials).

After these reaction time exclusions, we looked at pretest accuracy across all problem types, using proportion correct our accuracy measure. Mean proportion correct was 0.56 ($SD = 0.20$), and accuracy was somewhat normally distributed around this mean (see Figure 19). Given that participants who performed very well on the pretest had little to gain from the training videos, we excluded participants who performed better than one standard deviation above the mean – thus, we excluded an additional 45 participants whose proportion of problems solved correctly at pretest was greater than 0.76.

Based on this exclusion criteria, our final sample included 217 participants (matching – gesture principle: $n = 50$, matching – gesture algorithm: $n = 55$, mismatching – gesture principle: $n = 58$, mismatching – gesture algorithm: $n = 54$).

Figure 22. Histogram of proportion correct at pretest.



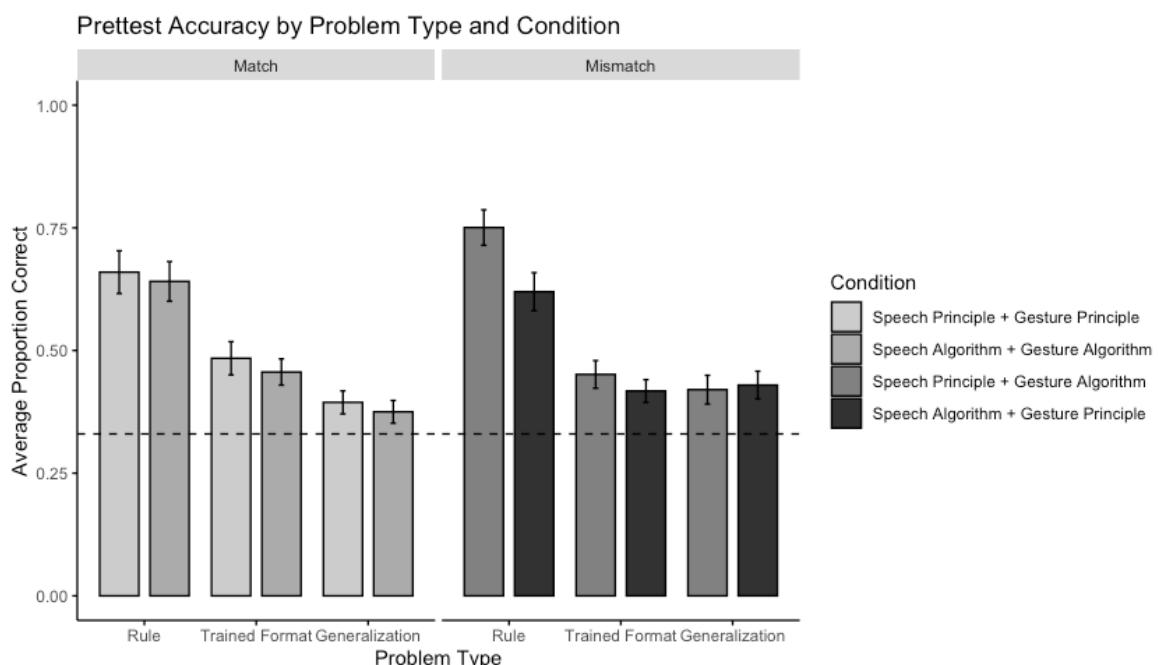
Pretest Accuracy: We first wanted to confirm that there were no significant differences on pretest accuracy by randomly-assigned condition. Because participants had not yet received any training instruction other than the rule videos (which were the same for all conditions), we

did not expect to see differences in pretest accuracy or reaction time on problem-solving by condition. To test this hypothesis, we ran a mixed-effects binomial logistic regression model with match status (matching or mismatch), gesture type (principle or algorithm), problem type (rule, trained format, or generalization), and interactions as main effects; reaction time as a covariate; participant as a random effect; and proportion correct on pretest items as the outcome variable. An analysis of variance of the model revealed no effects of match status ($\chi^2 (1) = 0.4, p = 0.6$) or gesture type ($\chi^2 (1) = 0.4, p = 0.5$), a significant effect of problem type ($\chi^2 (2) = 186.7, p < 0.000$), a trending interaction between match status and gesture type ($\chi^2 (2) = 2.9, p = 0.09$) and a significant interaction between match status and problem type ($\chi^2 (2) = 6.4, p = 0.04$). There was not a significant association between reaction time and pretest accuracy ($\beta = 0.002, z = 0.2, p = 0.9$). Participants performed better on rule problems compared to trained format ($\beta = 0.8, z = 4.4, p < 0.000$) and generalization problems ($\beta = 1.1, z = 6.5, p < 0.000$), and on trained format problems compared to generalization problems ($\beta = 0.4, z = 2.5, p = 0.01$).

To explore the trending interaction of match status and gesture type, and the interaction of match status and problem type, we used simpler mixed-effects binomial logistic regression models to predict accuracy on each problem type by match status, gesture type, and an interaction. Thus, we ran separate models of the same structure as the omnibus model for each problem type: rule problems, trained format problems, and generalization problems. All models had reaction time as a covariate, participant as a random effect, and proportion correct on pretest items as the outcome variable. For rule problems, there was a significant interaction between match status and gesture type, with participants in the mismatching – speech principle + gesture algorithm condition performing better than all other conditions ($\beta = 0.9, z = 2.0, p < 0.05$). There were no significant main effects of match status or gesture type nor interactions on accuracy on

trained format problems or generalization problems. The results of the pretest analysis suggested that random-assignment had not been completely successful in ensuring that participants did not differ by training condition in their ability to solve rule problems prior to beginning the training phase of the study. Thus, we controlled for pretest accuracy on rule problems in all subsequent analyses on training and posttest accuracy.

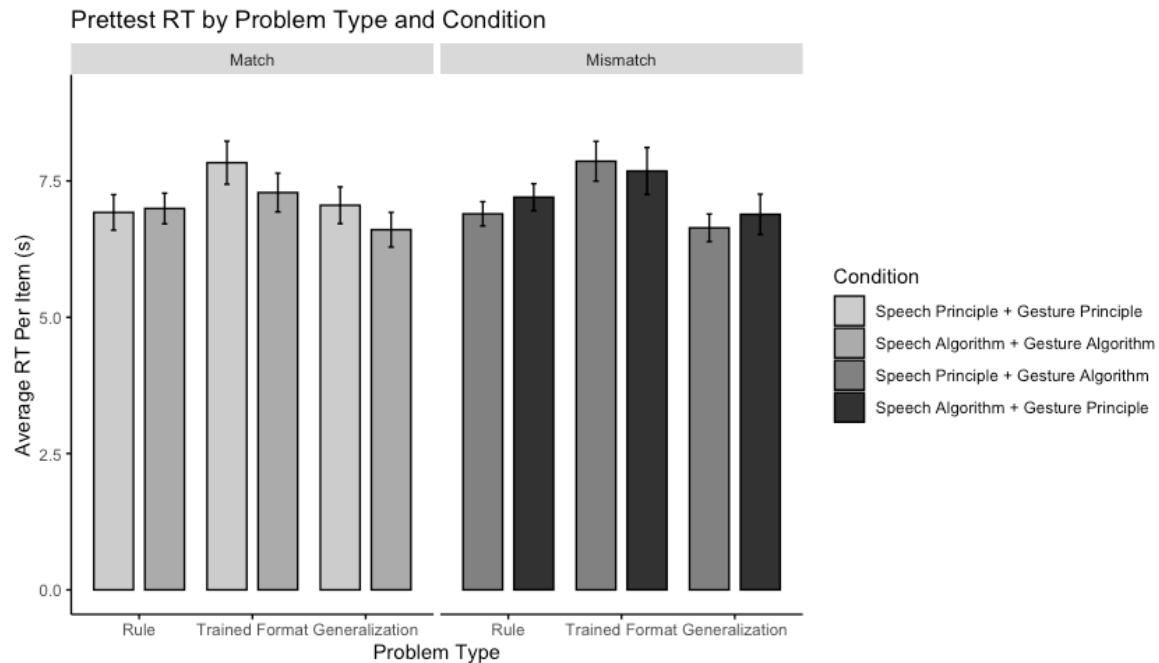
Figure 23. Pretest accuracy by problem type and condition.



Pretest Reaction Time: To explore whether participants responded more quickly on certain problem types and whether this varied by condition, we ran a generalized linear model with match status (matching or mismatch), gesture type (principle or algorithm), problem type (rule, trained format, or generalization), and interactions as main effects to predict mean reaction time. An analysis of variance of the model revealed no effect of match status ($\chi^2 (1) = 0.2, p = 0.7$) and gesture type ($\chi^2 (1) = 1.2, p < 0.000$), and a significant effect of problem type ($\chi^2 (1) = 14.9, p = 0.001$). Participants across conditions were significantly slower on trained format

problems than on rule problems ($\beta = 0.7$, $t = 2.8$, $p < 0.01$) and on generalization problems ($\beta = 0.9$, $t = 3.7$, $p < 0.001$). This analysis gave us confidence that participants did not differ by training condition in their problem-solving speed prior to beginning the training phase of the study.

Figure 24. Pretest reaction time by problem type and condition.



Training Accuracy: Prior to looking at accuracy on training, we first looked at reaction time for training trials, and excluded data for trials on which participants had a reaction time more than 3 standard deviations above the mean reaction time for this phase of the study, as well as trials for which reaction time was recorded as “0” seconds by Qualtrics, likely indicating that participants’ response time for these trials was below the sensitivity threshold for Qualtrics’ timing questions. The mean reaction time for training trials was 8.2 seconds ($SD = 8.3$ seconds), and thus, we excluded trials on which participants took more than 33.1 seconds to respond. This resulted in excluding 2.7% of training trials (46 trials).

To explore training accuracy by condition, we ran a mixed-effects binomial logistic regression model with match status (matching or mismatch), gesture type (principle or algorithm), and an interaction as main effects; reaction time and pretest accuracy on rule problems as covariates; participant as a random effect; and proportion correct on training items as the outcome variable. An analysis of variance of the model found no effects of match status ($\chi^2(1) = 0.01, p = 0.9$), gesture type ($\chi^2(1) = 0.08, p = 0.8$), nor an interaction ($\chi^2(1) = 1.1, p = 0.3$). Reaction time was not significantly associated with training accuracy ($\beta = 0.01, z = 0.3, p = 0.8$). Pretest accuracy on rule problems was significantly positively associated with training performance ($\beta = 1.1, z = 4.5, p < 0.000$).

Figure 25. Training accuracy by condition.

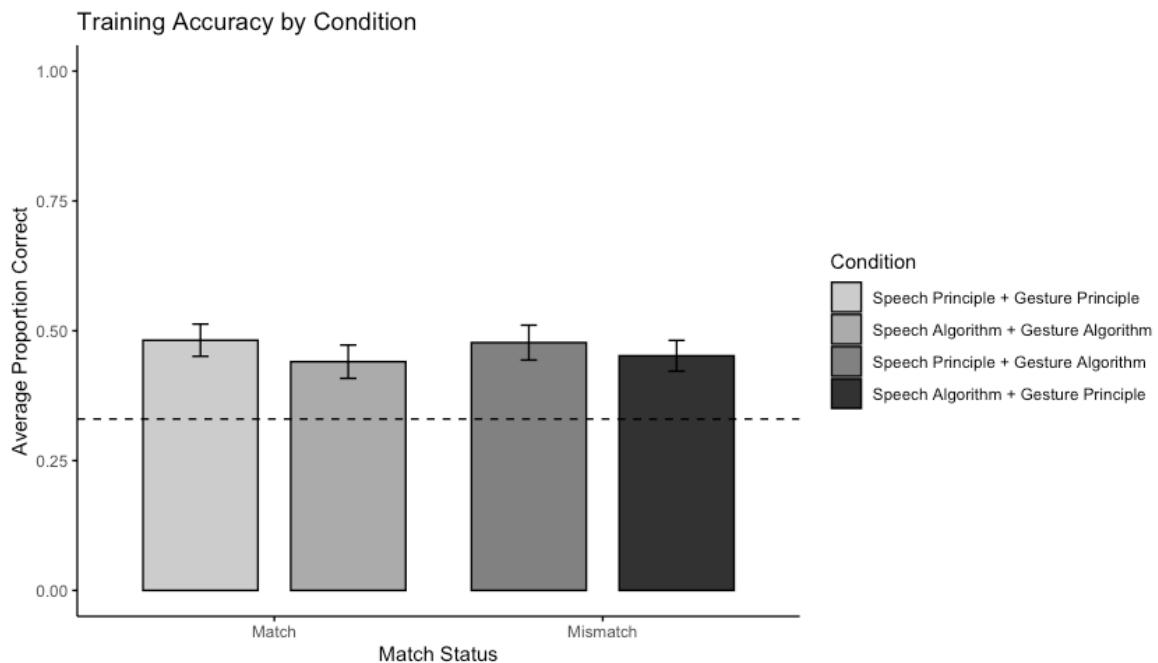
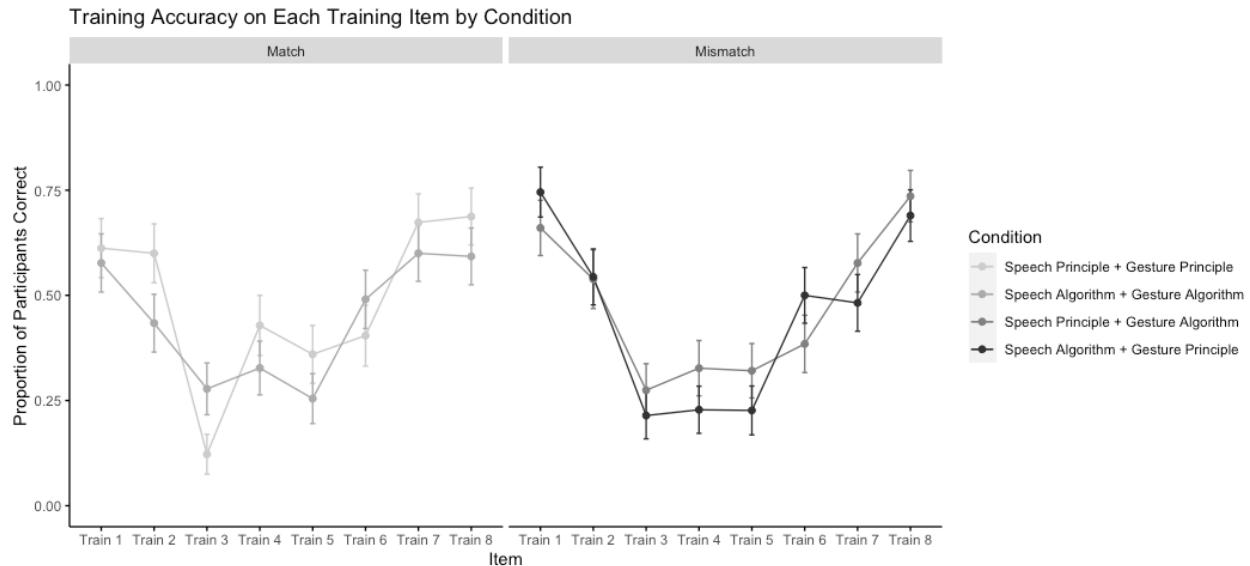


Figure 26. Training accuracy on each training trial by condition.

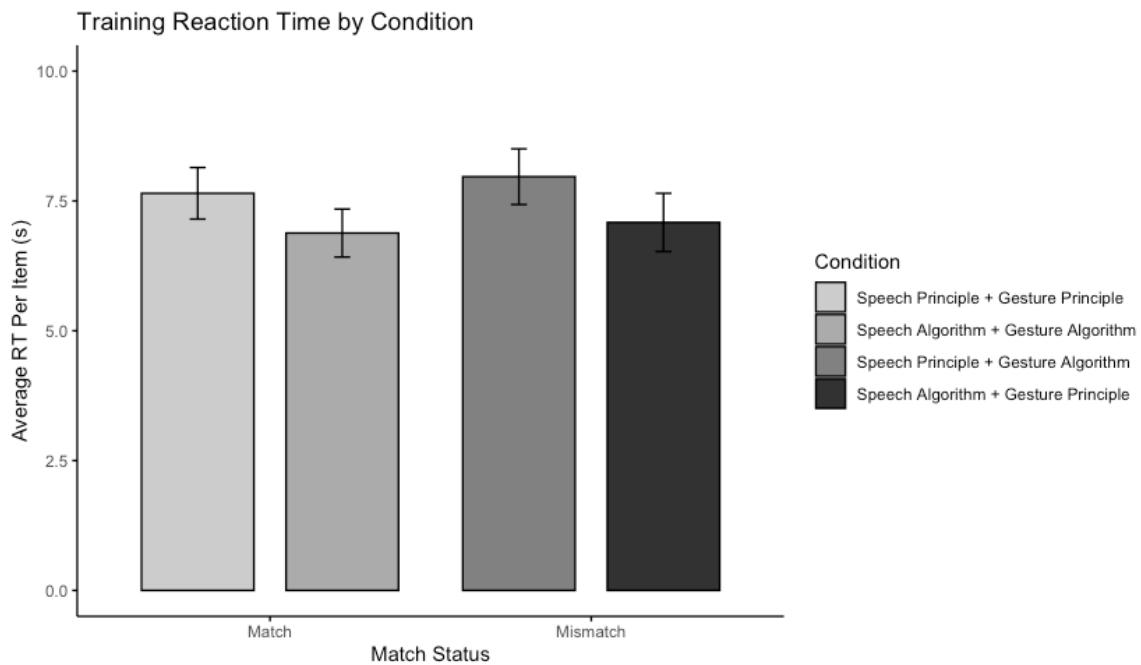


We also explored whether participants improved on training items across the 8 problems they solved during training, and whether this varied by condition. To test this, we ran another mixed-effects binomial logistic regression model with match status (matching or mismatch), gesture type (principle or algorithm), training item (items 1-8), and interactions as main effects; reaction time and pretest accuracy on rule problems as covariates; participant as a random effect; and accuracy on each item (0, 1) as the outcome variable. An analysis of variance of the model found no effect of match status ($\chi^2 (1) = 0.03, p = 0.9$) or gesture type ($\chi^2 (1) = 0.6, p = 0.4$), a significant effect of item ($\chi^2 (7) = 161.1, p < 0.000$), and no interactions. Reaction time was marginally negatively associated with training accuracy across items ($\beta = -0.1, z = -1.8, p = 0.08$) and accuracy on rule problems at pretest was significantly positively associated with training accuracy across items ($\beta = 0.4, z = 4.6, p < 0.000$). Across conditions, participants showed a u-shaped pattern of performance, performing well on the first two training items, struggling with items 3 and 4, and showing gradual improvement on the remaining items, returning to high

levels of performance on items 7 and 8. Items were equated for difficulty, so it is unclear why participants struggled with particular training items compared to others.

Training Reaction Time: To explore whether reaction time on training problems varied by condition, we ran a generalized linear model with match status (matching or mismatch), gesture type (principle or algorithm), and an interaction as main effects to predict reaction time. An analysis of variance of the model revealed no effects of match status ($\chi^2(1) = 0.3, p = 0.6$) or gesture type ($\chi^2(1) = 0.03, p = 0.9$), nor an interaction ($\chi^2(1) = 2.5, p = 0.1$),

Figure 27. Training reaction time by condition.



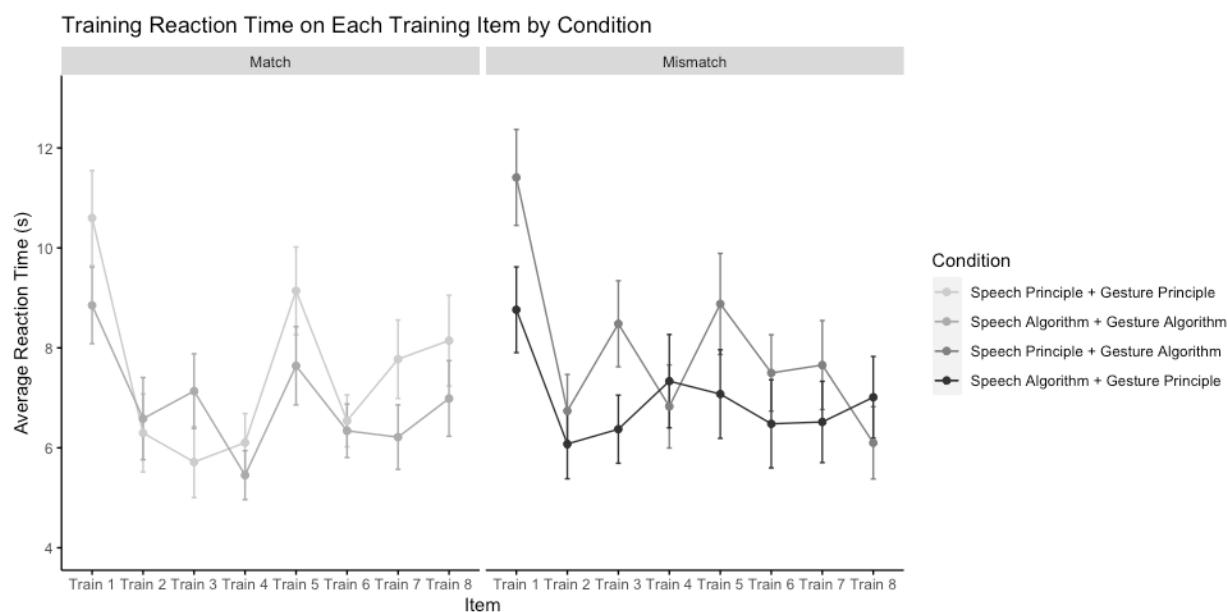
We also explored whether participants improved on reaction time across the 8 problems they solved during training, and whether this varied by condition. To test this, we ran another generalized linear model with match status (match or mismatch), gesture type (principle or algorithm), training item (items 1-8), and interactions as main effects to predict reaction time. An analysis of variance of the model revealed no effects of match status ($\chi^2(1) = 0.7, p = 0.4$) or gesture type ($\chi^2(1) = 0.5, p = 0.5$), a significant effect of item ($\chi^2(7) = 60.1, p < 0.000$), and a

significant interaction between match status and gesture type ($\chi^2 (1) = 8.4, p = 0.004$).

Participants were slower on item 1 compared to all other items, possibly as a result of acclimating to the training procedure of first watching a video, followed by problem-solving.

Participants in the mismatching – speech principle + gesture algorithm condition performed more slowly across training items compared to the other conditions ($\beta = 4.4, z = 2.8, p = 0.006$).

Figure 28. Training reaction time on each training trial by condition.



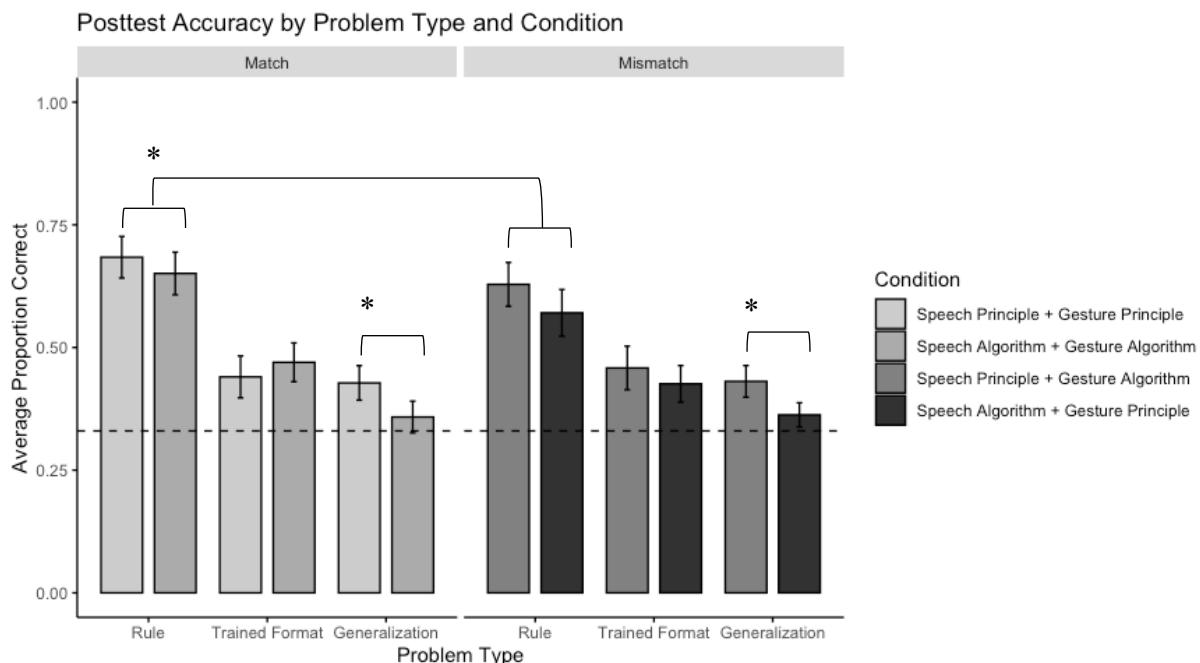
Posttest Accuracy: Prior to looking at accuracy at posttest, we first looked at reaction time for posttest items, and excluded data for trials on which participants had a reaction time more than 3 standard deviations above the mean reaction time for this phase of the study, as well as trials for which reaction time was recorded as “0” seconds by Qualtrics, likely indicating that participants’ response time for these trials was below the sensitivity threshold for Qualtrics’ timing questions. The mean reaction time for posttest items was 7.1 seconds ($SD = 5.7$ seconds), and thus, we excluded trials on which participants took more than 24.2 seconds to respond. This resulted in excluding 2.0% of posttest trials (92 trials).

To explore whether participants learned differentially depending on the instruction they received during training, we ran a mixed-effects binomial logistic regression model with match status (matching or mismatch), gesture type (principle or algorithm), problem type (rule, trained format, or generalization), and interactions as main effects; reaction time and pretest accuracy on rule problems as covariates; participant as a random effect; and proportion correct on posttest items as the outcome variable. An analysis of variance of the model revealed no significant effects of match status ($\chi^2 (1) = 0.6, p = 0.5$) or gesture type ($\chi^2 (1) = 0.2, p = 0.7$), a significant main effect of problem type ($\chi^2 (2) = 58.5, p < 0.000$), and a three-way interaction of match status, gesture type, and problem type ($\chi^2 (2) = 6.2, p < 0.05$). Reaction time was not significantly associated with posttest accuracy ($\beta = -0.01, z = -0.5, p = 0.6$) whereas accuracy on rule problems at pretest was significantly positively associated with posttest accuracy ($\beta = 1.7, z = 9.4, p < 0.000$). Participants performed significantly better on rule problems compared to training problems ($\beta = 0.9, z = 4.8, p < 0.000$) and generalization problems ($\beta = 0.8, z = 4.3, p < 0.000$).

To explore the interaction between match status, gesture type, and problem type, we used simpler mixed-effects binomial logistic regression models to predict accuracy on each problem type by match status, gesture type, and an interaction. Thus, we ran separate models of the same structure as the omnibus model for each problem type: rule problems, trained format problems, and generalization problems. All models had reaction time and accuracy on rule problems at pretest as covariates, participant as a random effect, and proportion correct on pretest items as the outcome variable. Participants in the matching conditions performed significantly better on rule problems compared to participants in the mismatching conditions ($\beta = 0.6, z = 2.5, p = 0.01$). For generalization problems, there was an interaction of mismatch status and gesture

type. For the mismatching conditions, participants who received speech principle + gesture algorithm performed better than participants who received speech algorithm + gesture principle ($\beta = 0.6$, $z = 2.4$, $p = 0.02$). This trend was opposite for the matching conditions: participants who received speech principle + gesture principle performed better than participants who received speech algorithm + gesture algorithm ($\beta = 0.6$, $z = 2.4$, $p = 0.02$).

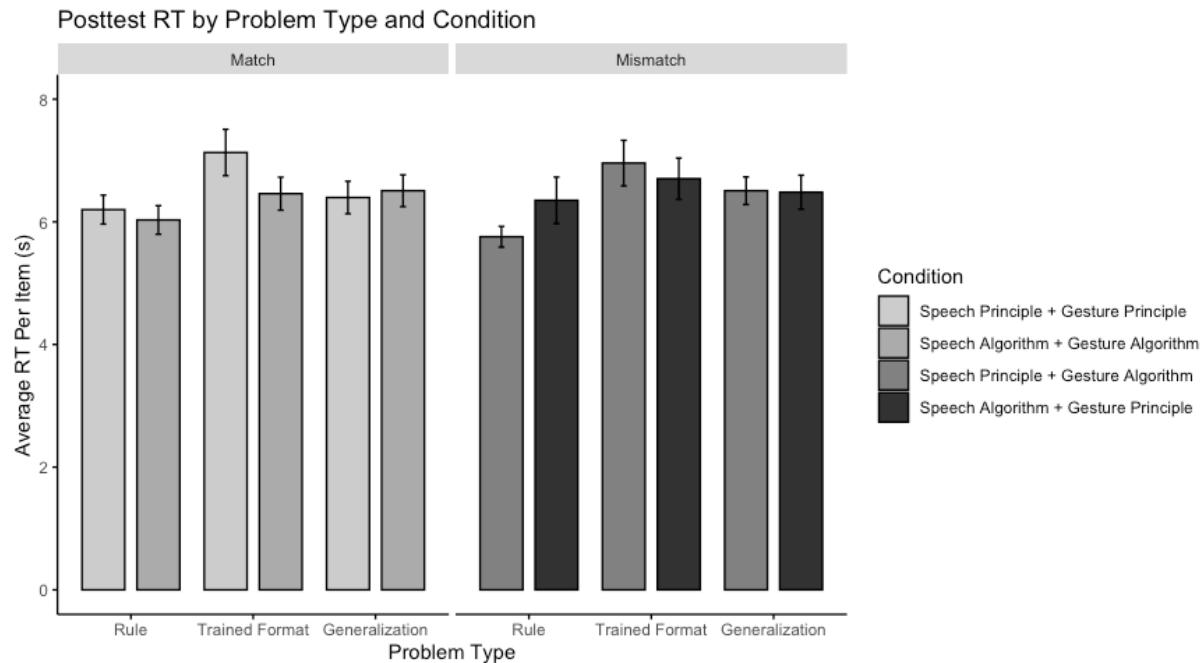
Figure 29. Posttest accuracy by problem type and condition.



Posttest Reaction Time: To explore whether participants responded more quickly on certain problem types and whether this varied by condition, we ran a generalized linear model with match status (matching or mismatch), gesture type (principle or algorithm), problem type (rule, trained format, or generalization), and interactions as main effects to predict reaction time on each trial. An analysis of variance of the model revealed no significant effects of match status ($\chi^2 (1) = 0.002$, $p = 1.0$) or gesture type ($\chi^2 (1) = 1.0$, $p = 0.3$), a significant effect of problem type ($\chi^2 (2) = 12.1$, $p = 0.002$), and no interactions. Participants were significantly faster on rule

problems compared to trained format problems ($\beta = -0.7$, $t = -3.5$, $p = 0.001$) and marginally faster on rule problems compared to generalization problems ($\beta = -0.4$, $t = -1.9$, $p = 0.06$).

Figure 30. Posttest reaction time by problem type and condition.

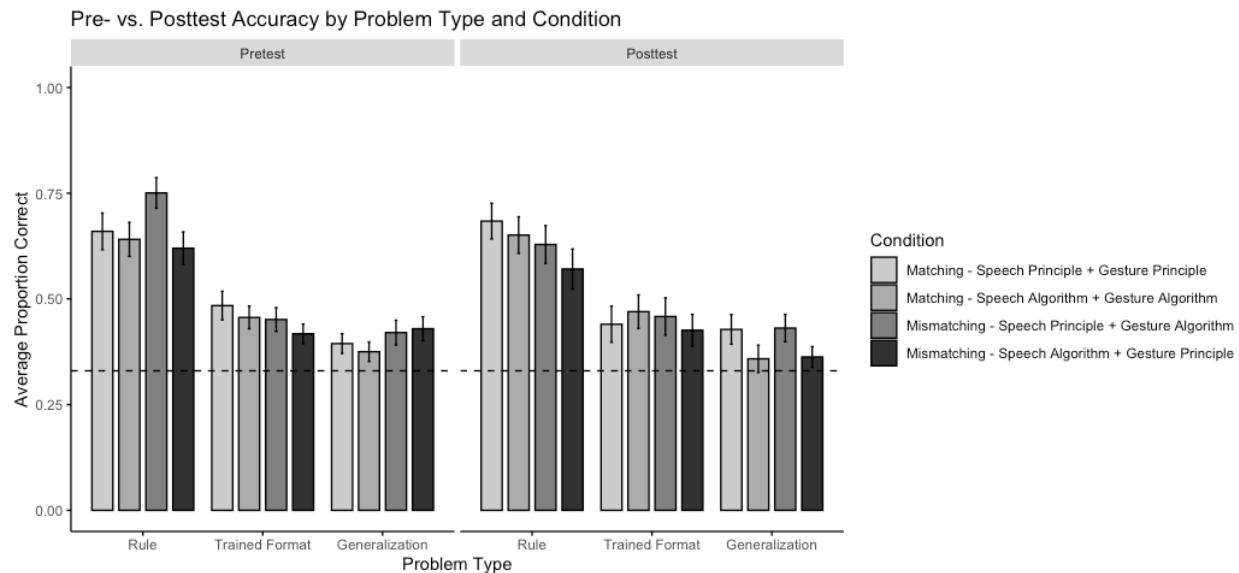


Comparing Pre- to Posttest Accuracy: To explore whether participants improved from pre- to posttest depending on the instruction they received during training, we ran a mixed-effects binomial logistic regression model with match status (matching or mismatch), gesture type (principle or algorithm), problem type (rule, trained format, or generalization), assessment (pretest or posttest), and interactions as main effects; reaction time as a covariate; participant as a random effect; and proportion correct at each assessment as the outcome variable. An analysis of variance of the model revealed a significant effect of problem type ($\chi^2 (2) = 349.5$, $p < 0.000$), a trending effect of trial type ($\chi^2 (1) = 3.4$, $p = 0.06$), a significant interaction of match status, problem type, and trial type ($\chi^2 (2) = 6.1$, $p < 0.05$), and a trending four-way interaction between

match status, gesture type, problem type, and assessment type ($\chi^2 (2) = 5.3, p = 0.07$). Reaction time was significantly negatively associated with accuracy ($\beta = -0.04, z = -2.8, p = 0.01$).

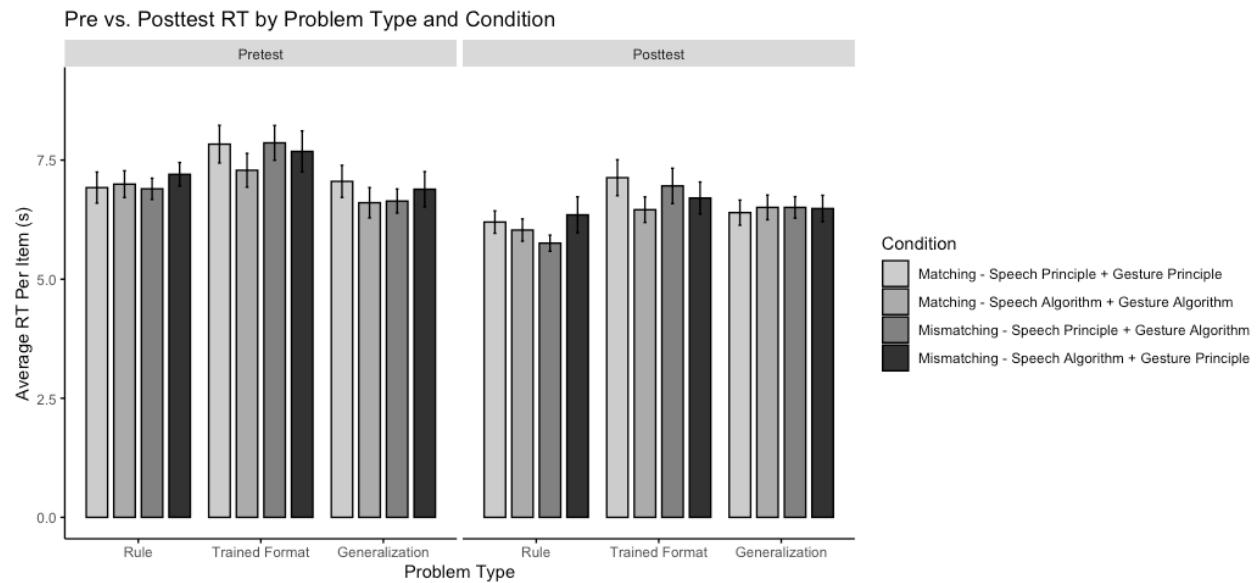
To explore the interaction between match status, gesture type, problem type, and assessment type, we used simpler models to predict accuracy on each problem type by match status, gesture type, assessment type, and interactions. Thus, we ran separate models of the same structure as the omnibus model for each problem type: rule, trained format, and generalization. Results indicated that the interactions of problem type, match status, and gesture type that we observed at posttest were driven both by improved performance from pre- to posttest in some conditions and decreased performance from pre- to posttest by other conditions. There was an interaction of match status and trial type for rule problems, with participants in the matching conditions performing significantly better from pre- to posttest ($\beta = 0.6, z = 2.8, p = 0.005$) and the mismatching conditions performing significantly worse from pre- to posttest ($\beta = -0.6, z = -2.8, p = 0.005$). There was a three-way interaction of match status, gesture type, and trial type for generalization problems, indicating that for the mismatching conditions, participants who received speech principle + gesture algorithm improved from pre- to posttest, whereas participants who received speech algorithm + gesture principle worsened from pre- to posttest ($\beta = 0.6, z = 2.4, p = 0.02$). This trend was opposite for the matching conditions: participants who received speech principle + gesture principle improved from pre- to posttest on generalization problems whereas participants who received speech algorithm + gesture algorithm worsened from pre- to posttest ($\beta = 0.6, z = 2.4, p = 0.02$).

Figure 31. Pretest vs. posttest accuracy by problem type and condition.



Comparing Pre- to Posttest Reaction Time: As a different measure of performance, we examined whether participants were able to solve problems at posttest more quickly compared to pretest. We ran a generalized linear model with condition (speech alone, gesture alone, or speech+gesture), problem type (rule, trained format, or generalization), assessment (pretest or posttest), and interactions as main effects and mean reaction time on each assessment item as the outcome variable. An analysis of variance of the model revealed significant effects of problem type ($\chi^2(2) = 22.8, p < 0.000$) and trial type ($\chi^2(1) = 30.1, p < 0.0000$). Participants in all conditions were significantly faster at posttest compared to pretest on rule problems ($\beta = -0.9, t = -4.8, p < 0.000$) and on trained format problems ($\beta = -0.9, t = -3.3, p < 0.001$).

Figure 32. Pretest vs. posttest reaction time by problem type and condition.



Discussion

Our results suggest that the type of nondeclarative knowledge depicted in gesture may impact learning in important ways, and that gesture may promote learning differently when it reiterates spoken knowledge than when it conveys knowledge that is distinct from spoken knowledge. Specifically, we observed differences in posttest accuracy on certain problem types based on the type of training instruction participants received. Participants in the matching conditions improved their knowledge of the rule structure from pre- to posttest compared to participants in the mismatching conditions, suggesting that gesture can benefit learning to some degree even when it reiterates spoken knowledge and does not provide complementary information. This finding is unexpected, as previous work suggests that gesture may be more beneficial for learning when it conveys information that differs from co-speech as opposed to information that matches co-speech (Hostetter, 2011).

For generalization problems, participants' performance varied depending on the type of strategy they saw conveyed in gesture, and whether or not this gesture matched the speech

strategy it accompanied. This result may suggest that the content of gesture impacts learning differently depending on the context in which it appears. However, an alternative, and more parsimonious, explanation for this result is that gesture did *not* play a role in shaping learners' knowledge, and rather, the *speech* strategy that participants received during training caused differences in learning by condition – specifically, participants who received a principle strategy in speech outperformed participants who received an algorithm strategy in speech on generalization problems. The principle strategy provides a general strategy that can be applied to all problem types, including generalization problems, whereas the algorithm grouping strategy (though it highlights and reinforces understanding of two-symbol rule combinations) provides a specific strategy for solving trained format problems. Indeed, Perry (1991) found that children who received instruction on math equivalence problems with a principle (equalizer) strategy were more likely to transfer their knowledge to novel problems than children who were instructed on an add-subtract (algorithmic) strategy. Our finding that adult participants benefited from the algorithm strategy when it was presented in gesture complements the result by Singer and Goldin-Meadow (2005) that children benefited from an algorithmic strategy when it was presented *uniquely* in gesture. Adult participants in the mismatching conditions performed like the children from Singer and Goldin-Meadow (2005) on generalization problems, benefiting from the algorithmic gesture only when it was presented in gesture. However, adults in the matching conditions succeeded on rule problems regardless of the strategy they received and succeeded on generalization problems when they received the principle strategy in speech. These findings suggest that adults may *not* benefit from multiple strategies simultaneously conveyed in speech+gesture in the same way that children do. This result thus poses an interesting

developmental question about how (and whether) speech and gesture differentially promote learning throughout the lifespan.

The lack of significant improvement from pre- to protest and the low performance overall on trained format problems during both training and posttest across conditions is somewhat surprising. Because participants received additional practice (with feedback) specifically on these problem types, we expected to see improved performance on these problems at posttest regardless of training condition. The finding that participants maintained their performance on trained format problems from pre- to posttest while increasing their problem-solving speed may signal a general practice effect, but it is also possible that participants improved their problem-solving speed as a result of insight into more efficient problem-solving strategies during training (even if they failed to apply such strategies effectively).

In future studies, it would be useful to have participants explain their problem-solving strategies after completing the posttest in order to better understand what type of information learners gained from the different training conditions, and how they applied this knowledge. It will also be important to investigate whether our findings replicate with a different algorithmic strategy, such as the add-subtract strategy used by Singer and Goldin-Meadow (2005). In contrast to the grouping strategy used in the present study, this algorithm can be applied to all problem structures (rather than to problems of one structure), and thus, may impact learning in different but meaningful ways. Furthermore, given that our study was inspired by Singer and Goldin-Meadow (2005), which focused on math equivalence learning with children, we should explore whether our findings replicate in a younger population, asking whether the *type* of strategy (principle vs. algorithm) conveyed in each modality (speech vs. gesture) impacts how children learn math equivalence. Finally, we were unable to collect a large amount of followup

data from participants in the present study. We provided participants with the option to complete a followup posttest 24-hours after the immediate posttest, but only 71 participants completed this followup, and of these participants, only 30 were included in our final sample for day 1 of the study, resulting in a followup sample that is far too small to glean meaningful findings from.

Given that the benefits of observing gesture on learning have been shown to emerge and persist over time in multiple previous studies (e.g., Cook, Duffy, & Fenn; 2013; Wakefield et al., 2019; Congdon et al., 2017) it is important that future research prioritizes collecting useful followup data in order to fully understand gesture's impact on learning. The 30 participants who completed the followup also answered a short questionnaire at the end of the followup and were asked if they noticed gestures in the videos they watched from the previous day. 11 of the 30 participants indicated that they noticed the gestures. Though this is a small representation of our sample, as observed in Study 2, this result may indicate that many participants failed to pay attention to the gestures during training or did not find them to be informative. As discussed in Chapter 3, given that participants completed the study at home on their personal computers, they may not have engaged with gesture instruction in the same manner that they would have had they completed the study in a laboratory or classroom setting. Differences in performance on generalization problems at posttest indicate that the *speech* strategy participants received during training impacted their ability to solve generalization problems while the gesture strategy did not, suggesting that at least some participants may have *listened* to the videos without *watching* them and seeing the accompanying gestures. However, performance on rule problems from pre- to posttest provides some evidence that participants did pay attention to the gestures they saw in the videos – specifically, participants in the matching conditions performed better than the mismatching conditions at posttest on *rule* problems, and this effect was driven by participants in

the matching conditions improving on rule problems from pre- to posttest, whereas participants in the mismatching conditions worsened from pre- to posttest on these problems. If participants were only listening to the speech, and not watching the gestures, performance should have been equal across conditions on rule problems, or only better for conditions with the same speech strategy. Future studies that use video stimuli in an online research paradigm can distinguish between these possibilities by including attention checks to ensure that participants both listen to and watch the video stimuli. Such work will also be important for understanding how gesture may translate as a learning device to online platforms.

CHAPTER FIVE: GENERAL DISCUSSION AND FUTURE DIRECTIONS

The experiments presented in this dissertation explore how gesture interacts with implicit and explicit learning paradigms, and how gesture may help learners bridge their non-declarative and declarative knowledge. The results reported demonstrate that gesture likely benefits learning through complex mechanisms, and extends previous research on gesture learning in children to adult learners, thus adding to a growing body of work showing that gesture instruction can promote learning throughout the lifespan.

In Study 1, we successfully adapted implicit and explicit learning paradigms from the cognitive literature for training of a novel concept, and we incorporated gesture into this instruction. During training, participants were shown many examples of correctly-solved equations, generated by a symbolic math system. These training conditions were identical except for the *instructions* participants were given at the start of training: one group was told to intentionally try to learn the rules underlying the equations (explicit learning), one group was told simply to remember each equation (implicit learning), and a third “hybrid” group was told to remember each equation for the first half of training and to intentionally learn the rules for the second half (implicit→explicit learning). Half the participants in each condition also saw gestures during the videos they watched. Participants in all training conditions identified correctly-solved novel equations above chance levels after viewing many videos of equations during training, but participants who received some explicit instruction during training (explicit learning or implicit→explicit learning) showed higher levels of learning across different problem types than participants who only received implicit instruction. Training conditions also varied in their reaction time during training and posttest; during both phases of the study, the explicit conditions performed more slowly than the implicit conditions, and the implicit→explicit

conditions performed more quickly than the other two learning conditions. Importantly, these differences in reaction time did not account for differences in posttest accuracy across conditions – though reaction time was significantly positively associated with accuracy (indicating that more time spent problem-solving led to better accuracy), accuracy differences by condition still emerged even when controlling for reaction time during posttest and training, indicating that the mechanisms by which explicit and implicit→explicit instruction promote better learning than implicit instruction are not solely based on encouraging learners to spend more time solving problems (in fact, the implicit→explicit condition demonstrated better accuracy than the implicit condition while simultaneously showing faster reaction times than both the implicit and explicit conditions). Thus, Study 1 represents an important replication and extension of previous experiments on implicit vs. explicit learning of artificial grammatical rules (e.g., Reber, 1976; Reber et al., 1980; Mathews et al., 1989) to learning of a *mathematical* rule structure. These findings add to a growing body of work suggesting that implicit and explicit learning processes are active across different domains of knowledge acquisition (e.g., Batterink, Paller, & Reber, 2019).

Study 1 also explored how *gesture* interacts with implicit and explicit learning processes. Gestural instruction has been shown to impact learning in ways that mirror both implicit and explicit learning. Just as non-declarative knowledge is acquired through implicit learning mechanisms, information conveyed in gesture can be learned unconsciously (e.g., Ellis Weismer & Hesketh, 1993), is rooted in motor experience (e.g., Ping, Goldin-Meadow, & Beilock, 2014), and occurs gradually over time (e.g., Cook, Duffy, & Fenn, 2013). Gesture also integrates seamlessly with speech (e.g., McNeill, 1992), a vehicle for declarative knowledge, and promotes generalization of learned knowledge (e.g., Congdon et al., 2017; Wakefield et al., 2018a). Both

the ability to verbalize and generalize a learned concept signify that explicit learning has occurred (e.g., Eichenbaum & Cohen, 2001). Study 1 is the first of its kind to directly examine how gesture interacts with traditional implicit vs. explicit learning paradigms, and thus bring together the cognitive and gesture-learning literatures. Posttest accuracy by participants in the gesture training conditions on problems designed to precisely test *declarative knowledge* of the rules underlying the symbolic system (rather than ability to *apply* these rules) showed that gesture training only benefited performance on these problems when it was presented in either an explicit or implicit→explicit learning context. Despite above-chance performance on novel problems at posttest of the same format as problems seen during training, participants in the implicit and implicit+gesture conditions did not perform above chance at identifying novel valid equations depicting *rules* underlying the mathematical system. This finding suggests that gesture may be particularly helpful for learning when participants *intend* to learn a concept, and provides evidence in favor of our original hypothesis: if gesture conveys non-declarative knowledge, gesture may not promote learning in an implicit context, but may be effective in an explicit context, as the latter allows learners to simultaneously benefit from the non-declarative knowledge depicted in gesture while engaging in explicit learning processes. This result is not incompatible with findings from previous work showing that gestural information can be learned unconsciously, but it suggests that the context in which gesture is presented may impact the degree to which gesture benefits learning and the *type* of knowledge that is gained.

Furthermore, in looking at the accuracy trends on a followup test completed by about one-third of all Study 1 participants 24-hours after the immediate posttest, the explicit+gesture and implicit→explicit+gesture conditions appear to maintain their superior performance on rule problems (though this trend is more pronounced in the explicit+gesture condition). Interestingly,

the implicit+gesture conditions show a trending advantage over their implicit alone counterparts on the followup test on trained problems and rule problems, replicating previous findings demonstrating that benefits of gestural instruction can emerge over time (e.g., Cook, Duffy, and Fenn; 2013; Wakefield et al., 2019a; Wakefield et al., 2019; Congdon et al., 2017). These findings could be explored in a future study specifically aimed at tracking the effects of gesture with implicit vs. explicit instruction over time. With a larger sample, researchers could also test whether learners' ability to articulate the rules emerges over time, and whether this varies depending on the initial instruction received. If the accuracy trends observed in our small followup sample are shown to be robust, this would suggest that gesture can positively impact learning in all contexts, but that its benefits are delayed when it is presented in a solely implicit context. If gesture instruction is shown to increase the likelihood that participants can verbalize their knowledge of the rules, even when gesture is presented in an implicit learning context, this would provide important evidence that gesture cannot only benefit learning in different contexts, but that it actively plays a role in helping learners transition from a state of non-declarative to declarative knowledge.

Studies 2 and 3 explored how unspoken, non-declarative knowledge conveyed by gesture impacts learning depending on whether the gesture is paired with declarative spoken knowledge as well as on the type of unspoken knowledge conveyed by the gesture. In both studies, all participants were told they would be learning how to solve problems in a new symbolic math system. In Study 2, we created instructional conditions that manipulated whether gesture was presented alone or in conjunction with speech, and compared training in these conditions to instruction with speech alone. The results of Study 2 were inconclusive, as participants across conditions showed poor performance on a posttest designed to measure their learning. Though

the gesture alone condition was experimental, we expected the speech + gesture condition to outperform the speech alone condition based on the vast amount of work demonstrating this pattern of results (e.g., Singer & Goldin-Meadow, 2005; Cook & Goldin-Meadow, 2006; Ping & Goldin-Meadow, 2008; Valenzano, Alibali, & Klatzky, 2003; Wakefield et al., 2018b; Congdon et al., 2017). Despite its lack of findings, Study 2 represents an important attempt to explore how gesture may (or may not) function as a learning device in the absence of meaningful speech – that is, when it is ungrounded in declarative knowledge.

The results of Study 3 were clearer than those of Study 2, but still require further exploration. This study used the same design as Study 2, but all participants received speech + gesture instruction and the *type* of information conveyed by gesture in each condition (either a principled or algorithmic problem-solving strategy) as well as whether the gesture matched or mismatched its accompanying speech, varied by condition. Participants who saw gesture-speech matches during instruction performed better on rule problems at posttest compared to participants who saw gesture-speech mismatches, regardless of the type of strategy they saw. On problems designed to test learners' ability to generalize these rules to a novel problem structure, only participants who received a principle (equalizer) strategy in speech showed learning benefits, regardless of whether the gesture matched (equalizer) or differed from (grouping) this speech strategy. These findings suggest that gesture may impact learning in nuanced ways, and particularly with adult learners, it may not play as significant a role in learning as speech does. In a context where adult learners are prepared to learn a novel concept, gesture that *reinforces speech* can benefit learning of an underlying rule structure regardless of the strategy presented; in learning to generalize acquired knowledge to different problem structures, gesture does not

hinder learning when it presents information that *differs from speech*, as long as it is grounded in a *principle* speech strategy.

Though Studies 2 and 3 were unsuccessful, they represent important first steps in researching how observed gesture can promote learning in adult participants. The symbolic math task that participants were taught in Studies 1, 2, and 3 has been used in a previous study by Cook & Fenn (in prep.). This study found that instruction with gesture promoted better learning than instruction with speech alone for adult learners, but this effect did not emerge until 24-hours later, on a followup test. We did not observe benefits of gesture on immediate posttests for Studies 2 and 3, and were unable to collect followup data for the vast majority of participants, and thus, it is unclear whether any benefits of gesture in our learning paradigm may have emerged over time. Though we provided all participants who completed these studies with the opportunity to complete a followup test by messaging them directly through Amazon Mechanical Turk, very few participants took this opportunity. Even in Study 1, only about one-third of participants completed the followup tasks. High attrition rates across all studies indicate that perhaps we did not offer a strong enough incentive to participants for completing the followup study, or that the method provided by MTurk to contact participants (by sending them a message via the site's "bonus" system) is ineffective. In future studies, it will be important that the study protocol emphasize to participants the importance of completing the followup assessment and that all participants have easy access to this assessment (e.g., for an online study similar to ours, the study should be run on a site that allows researchers to collect participants' e-mail addresses in order to send them direct links to the followup survey, or on a site that otherwise offers a more researcher-friendly format for conducting longitudinal research; Prolific.co, a UK-based research site first launched in 2014, appears to be one such promising alternative to MTurk). Our

followup assessment also included questions asking participants whether they noticed gestures during training and if so, whether these gestures helped them solve the problems and how.

Having a large sample of responses to these questions could provide insight into what participants may have found useful or confusing about gestural instruction in our paradigm.

Other differences between the protocols of Studies 2 and 3 and that of Cook & Fenn (in prep.) regarding how participants were introduced to the rules of the symbolic math system, and the use of a pretest, may have impacted the results of Studies 2 and 3. Rather than introducing participants to the rules via instructional videos, Cook & Fenn (in prep.) had participants read basic descriptions of how the symbols could be combined and then tested participants on their knowledge of the rules. Participants were unable to continue with the experiment until they demonstrated understanding of the rules. This design by Cook & Fenn (in prep.) may have helped ensure that participants had a similar level of knowledge of the rules prior to receiving problem-solving instruction. Pretest performance in Studies 2 and 3 indicated that participants had varying levels of knowledge after watching the introductory rule videos, and though performance at pretest did not vary by randomly-assigned condition prior to training, poor performance across conditions in both studies may indicate that participants needed a stronger foundational knowledge of the rules in order to benefit from training. In contrast, Cook & Fenn (in prep.) did not include a pretest in their study design, and thus, we do not know from their results the degree to which the instructional conditions benefited learning. In Study 2, participants worsened from pre-to-posttest, and in Study 3, participants in some conditions showed only small benefits from training. Still, it is possible that being asked to solve trained format and generalization problems at pretest prior to receiving training may have been cognitively fatiguing for participants, such that they were less engaged when viewing the training

videos than they may have been had they not completed a pretest. Though the goal of including a pretest in our study design was to measure participants' baseline knowledge prior to training, this design may have drawbacks that need to be weighed carefully with potential benefits. For a future study protocol, a "compromise" may be to have participants reach a rule-learning criterion (as in Cook & Fenn, in prep.) and then solve a *small* number of trained format and generalization problems to serve as a measure of their baseline ability to apply these rules. Such a design would provide participants with equal grasp of the rules prior to training, while still providing researchers with a baseline measure of performance for comparison to posttest performance.

At first glance, the results of Study 1 and the lack of results of Study 2 may seem contradictory considering that both studies paired gesture with meaningless speech in some training conditions, and participants in Study 1 were able to benefit from gesture paired with this vacuous speech while participants in Study 2 were not. However, these study designs differed in important ways. First off, in Study 1, participants were told to *learn* the rules of the symbolic math system or to simply remember equations generated by the rules of the system whereas in Study 2, participants were *given* the rules at the beginning of the study and then taught problem-solving strategies for solving equations in the system. Because participants in Study 1 had no declarative knowledge (i.e., of the rules) to rely upon during training, they may have gleaned more from the non-declarative knowledge presented in gesture than participants in Study 2, who may have been less motivated to attend to the instruction after learning the rules necessary to solve all equations. Additionally, in Study 1, participants saw a grouping (algorithmic) gesture and in Study 2, participants saw an equalizer (principle) gesture. The results of Study 3 suggest that learners may gain different levels of understanding of the concept to be learned depending on the strategy presented in speech, and that gesture can promote learning in adults when it

reinforces speech. Again, because participants in Study 1 had no prior knowledge of the rules to rely upon, they may have benefited from the algorithmic gesture’s highlighting of two-symbol combinations that made up these rules. To explore this possibility, we could replicate both Study 1 and Study 2 but swap the gesture strategies used in each study (principle for Study 1 and algorithm for Study 2). These experiments would further specify which aspects of certain gestures promote learning, and in which contexts. Additionally, the participants in Study 1 had 96 training trials on which to observe and potentially benefit from the gesture strategy, whereas participants in Studies 2 and 3 had only 8 training trials. Because participants completed all study tasks online via their home computers, participants in all studies may have engaged with the study differently than if they had been in a laboratory or classroom setting. If participants were distracted or failed to look at the stimuli for any of the trials, this may have impacted their ability to learn from the gestures shown in the videos. In Study 1, participants had many “chances” to see gesture, and thus may have benefited from gesture even if they were distracted on *some* training trials; in Studies 2 and 3, however, *any* training trials on which participants were distracted were important learning opportunities lost. For future online research using video stimuli, we need to emphasize to participants the importance of paying attention to the videos, and perhaps include frequent attention checks to ensure that participants engage with the videos for the entirety of the study. Lastly, the posttest in Study 1 asked participants to identify correctly-solved equations rather than solve equations on their own (as in Studies 2 and 3). This method of measuring learning in Study 1 was based on posttests used in artificial grammar learning studies, whereas the posttest structure used in Study 2 was modeled from posttests used in math equivalence studies with children in which learners are asked to solve problems of different structures than ones seen during training. Even though participants were given the rules

in Studies 2 and 3, *generating* correct solutions to problems is a more challenging task than *recognizing* correct solutions, and we may have seen higher rates of learning in Studies 2 and 3 (and more meaningful differences between training conditions) had we used a testing method more similar to that used in Study 1. It is also important to acknowledge that the unprecedented times during which these data were collected may have impacted the results – all three studies were completed online during the Covid-19 pandemic, but Study 1 was completed in March of 2020 at the start of the pandemic, whereas Studies 2 and 3 were completed a few months later in June of 2020. Though this timeline may be negligible under ordinary circumstances, many popular websites for researchers, including Amazon Mechanical Turk, have seen an influx of new online workers as unemployment has risen nationwide (Schwartz, 2020; Simonite, 2020). While the participant pool on MTurk has remained diverse (Moss et al., 2020), many new workers may still be adjusting to online work and specifically the MTurk platform, resulting in lower performance than might be expected from more experienced workers.

Overall, the findings from Study 1 presented in this dissertation provide evidence that part of the mechanism through which gesture benefits learning is by serving as a vehicle for non-declarative knowledge. Though the results of Studies 2 and 3 were largely inconclusive, they offer some insight into how future work may better explore how the specific types of non-declarative knowledge conveyed in gesture may impact learning. Study 1 demonstrates that gesture can impact learning differentially depending on the context in which it is presented – particularly, gesture may promote more robust learning when observers are intending to learn a concept. These studies offer an important contribution to the gesture learning literature by showing how gesture instruction promotes understanding of a mathematical concept in *adult* learners. A metanalysis of gesture-learning studies by Hostetter (2011) found that children

benefit more from gesture instruction than do adults, and while this may be true, the current work demonstrates that gesture is still a useful learning device across the lifespan, even if it functions in complex ways depending on the learning context and the information to be learned. Furthermore, all three studies were designed to be conducted online with a large and diverse sample as a result of research constraints due to the Covid-19 crisis, and thus show that with methodological creativity, effective research need not be limited to in-person study designs.

Future work can use different methods in conjunction with the behavioral methods used in this dissertation to further specify how gesture facilitates learning in implicit vs. explicit learning contexts across the lifespan, and how non-declarative knowledge conveyed in different gestural strategies may promote different learning trajectories. For example, eye tracking technology could measure exactly what aspects of gesture learners attend to during implicit vs. explicit instruction, or during instruction with gesture that conveys a principle vs. algorithmic problem-solving strategy. These results would shed light on how gesture may guide learners' attention to relevant stimuli in different contexts. Additionally, future studies can use neuroimaging to explore how knowledge acquired through different types of gestural instruction is represented in the brain, and whether these representations vary depending on the context in which gesture is observed. Neuroimaging data could yield insight into how brain structures involved in implicit and explicit learning, as well as storage of non-declarative vs. declarative knowledge, may overlap with areas involved in gesture processing. Such information would help researchers and educators understand how information learned through gesture may transform a learner's understanding of a concept over time, perhaps by helping a learner gradually transition from a state of non-declarative to declarative knowledge.

REFERENCES

- Aleven, V. A., & Koedinger, K. R. (2002). An effective metacognitive strategy: learning by doing and explaining with a computer-based cognitive tutor. *Cognitive Science*, 26, 147-179.
- Alibali, M., Flevares, L., & Goldin-Meadow, S. (1997). Assessing knowledge conveyed in gesture: Do teachers have the upper hand? *Journal of Educational Psychology*, 89, 183-193.
- Altmann, G. T. M., Dienes, Z. and Goode, A. (1995) Modality independence of implicitly learned grammatical knowledge. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 21, 899-912.
- Batterink, L. J., Paller, K. A., & Reber, P. J. (2019). Understanding the Neural Bases of Implicit and Statistical Learning. *Topics in Cognitive Science*, 11, 1-22.
- Bayley, P. J., & Squire, L. R. (2002). Medial temporal lobe amnesia: Gradual acquisition of factual information by nondeclarative memory. *The Journal of Neuroscience*, 22, 5741-5748.
- Beilock, S. L. & Goldin-Meadow, S. (2010). Gesture changes thought by grounding it in action. *Psychological Science*, 21, 1605-1610.
- Bernardis, P., & Gentilucci, M. (2006). Speech and gesture share the same communication system. *Neuropsychologia*, 44, 178-190.
- Berry, D. C. (1994). Implicit learning: Twenty-five years on: A tutorial. In Carlo Umiltà & Morris Moscovitch (eds.), *Attention and Performance XV: Conscious and Nonconscious Information Processing*, 755-782. Cambridge, MA: MIT Press.
- Berry, D. C., & Broadbent, D. E. (1988). Interactive tasks and the implicit-explicit distinction. *British journal of Psychology*, 79, 251-272.
- Berry, D. C., & Broadbent, D. E. (1984). On the relationship between task performance and associated verbalizable knowledge. *Quarterly Journal of Experimental Psychology*, 36A, 209-231.
- Bisanz, J., & LeFevre, J. A. (1992). Understanding elementary mathematics. In J. D. Campbell (Ed.), *The nature and origins of mathematical skills* (pp. 113-136). Oxford, England: North-Holland.
- Bowers, K. S., Regehr, G., Balthazard, C., & Parker, K. (1990). Intuition in the context of discovery. *Cognitive Psychology*, 22, 72-110.

- Brinol, P., Petty, R. E., & Wheeler, S. C. (2006). Discrepancies between explicit and implicit self-concepts: Consequences for information processing. *Journal of Personality and Social Psychology, 91*, 154-170.
- Broadbent, D. E., Fitzgerald, P., & Broadbent, M. H. P. (1986). Implicit and explicit knowledge in the control of complex systems. *British Journal of Psychology, 77*, 33-50.
- Broaders, S. C., & Goldin-Meadow, S. (2010). Truth Is at Hand How Gesture Adds Information During Investigative Interviews. *Psychological Science, 21*, 623-628.
- Brooks, L. R. (1978). Non-analytic concept formation and memory for instances. In E. Rosch & B. B. Lloyd (Eds.), *Cognition and categorization*. Hillsdale, N.J.: Erlbaum, 1978.
- Brooks, L. R., & Vokey, J. R. (1991). Abstract analogies and abstracted grammars: Comments on Reber (1989) and Mathews et al. (1989). *Journal of Experimental Psychology: General, 120*, 316-323.
- Capone, N. C., & McGregor, K. K. (2005). The effect of semantic representation on toddlers' word retrieval. *Journal of Speech, Language, and Hearing Research, 48*, 1468-1480.
- Church, R. B., Ayman-Nolley, S., & Mahootian, S. (2004). The effects of gestural instruction on bilingual children. *International Journal of Bilingual Education and Bilingualism, 7*, 303-319.
- Church, R. B., Kelly, S., & Holcombe, D. (2014). Temporal synchrony between speech, action and gesture during language production. *Language, Cognition and Neuroscience, 29*, 345-354.
- Clayton, N. S., & Dickinson, A. (1998). Episodic-like memory during cache recovery by scrub jays. *Nature, 395*, 272-274.
- Cleeremans, A., Destrebecqz, A. & Boyer, M. (1998). Implicit learning: News from the front. *Trends in Cognitive Science, 2*, 406-416.
- Cohen, N. J., & Eichenbaum, H. (1993). *Memory, amnesia and the hippocampal system*. Cambridge, M.A.: MIT Press.
- Cohen, N. J., Eichenbaum, H., Deacedo, B., & Corkin, S. (1985). Different memory systems underlying acquisition of procedural and declarative knowledge. *Annals New York Academy of Sciences, 444*, 54-71.
- Cohen, A., Ivry, R. I., & Keele, S. W. (1990). Attention and structure in sequence learning. *Journal of Experimental Psychology: Learning, Memory and Cognition, 16*, 17-30.
- Cohen, N. J., & Squire, L. R. (1980). Preserved learning and retention of pattern-analyzing skill in amnesia: Dissociation of knowing how and knowing that. *Science, 210*, 207.

- Congdon, E. L., Novack, M. A., Brooks, N., Hemani-Lopez, N., O'Keefe, L., & Goldin-Meadow, S. (2017). Better together: Simultaneous presentation of speech and gesture in math instruction supports generalization and retention. *Learning and Instruction*, 50, 65-74.
- Cook, S. W., Duffy, R. G., & Fenn, K. M. (2013). Consolidation and transfer of learning after observing hand gesture. *Child development*, 84, 1863-1871.
- Cook, S. W., Mitchell, Z., & Goldin-Meadow, S. (2008). Gesture makes learning last, *Cognition*, 106, 1047-1058.
- 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, 211-232.
- Cook, S. W. & Tanenhaus, M. K. (2009). Embodied understanding: Speakers' gestures affect listeners' actions. *Cognition*, 113, 98-104.
- Corkin, S. (1968). Acquisition of motor skill after bilateral medial temporal-lobe excision. *Neuropsychologia*, 6, 255-265.
- Destrebecqz, A., & Cleeremans, A. (2001). Can sequence learning be implicit? New evidence with the process dissociation procedure. *Psychonomic Bulletin & Review*, 8, 343-350.
- Dienes, Z., & Berry, D. (1997). Implicit learning: below the subjective threshold. *Psychonomic Bulletin and Review*, 4, 3-23.
- Dienes, Z., Broadbent, D. E., & Berry, D. C. (1991). Implicit and explicit knowledge bases in artificial grammar learning. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, 17, 875-887.
- Dulany, D. E., Carlson, R. A., & Dewey, G. I. (1984). A case of syntactical learning and judgment: How conscious and how abstract? *Journal of Experimental Psychology: General*, 113, 541-555.
- Eichenbaum, H., & Cohen, N. J. (2001). *From conditioning to conscious recollection: Multiple memory systems of the brain*. New York: Oxford University Press.
- Ellis Weismar, S., & Hesketh, L. J. (1993). The influence of prosodic and gestural cues on novel word acquisition by children with specific language impairment. *Journal of Speech and Hearing Research*, 36, 1013-1025.
- Flevares, L. M. and Perry, M. 2001. How many do you see? The use of nonspoken representations in first-grade mathematics lessons. *Journal of Educational Psychology*, 93, 330-345.

- Fortin, N. J., Wright, S. P., & Eichenbaum, H. (2004). Recollection-like memory retrieval in rats is dependent on the hippocampus. *Nature*, 431, 188-190.
- Frensch, P. A., Buchner, A., & Lin, J. (1994) Implicit learning of unique and ambiguous serial transitions in the presence and absence of a distractor task. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 20, 567-584.
- Gabrieli, J. D. E. (1998). Cognitive neuroscience of human memory. *Annual Review of Psychology*, 49, 87-115.
- Garber, P., Alibali, M. W., & Goldin-Meadow, S. (1998). Knowledge conveyed in gesture is not tied to the hands. *Child Development*, 69, 75-84.
- Gluck, M. A. & Bower, G. H. (1988). From conditioning to category learning: An adaptive network model. *Journal of Experimental Psychology: General*, 117, 227-247.
- Gluck, M. A., Shohamy, D., & Myers, C. E. (2002). How do people solve the “weather prediction” task?: Individual variability in strategies for probabilistic category learning. *Learning and Memory*, 9, 408-418.
- Goldin-Meadow, S., Alibali, M. W., & Church, R. B. (1993). Transitions in concept acquisition: Using the hand to read the mind. *Psychological Review*, 100, 279-297.
- Goldin-Meadow, S., Cook, S. W., & Mitchell, Z. A. (2009). Gesturing gives children new ideas about math. *Psychological Science*, 20, 267-272.
- Gomez, R. L., & Schvaneveldt, R. W. (1994). What is learned from artificial grammars? Transfer tests of simple association. *Journal of Experimental Psychology: Learning, Memory and Cognition*. 20, 396-410.
- Goodrich, W., & Hudson Kam, C. L. (2009). Co-speech gesture as input in verb learning. *Developmental Science*, 12, 81-87.
- Goshen-Gottstein, Y., Moscovitch, M., & Melo, B. (2000). Intact implicit memory for newly formed verbal associations in amnesic patients following single study trials. *Neuropsychology*, 15, 570-578.
- Graf, P., & Schacter, D. L. (1985). Implicit and explicit memory for new associations in normal and amnesic subjects. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 11, 501-518.
- Greeno, J., & Riley, M. (1987). Processes and development of understanding. In F. E. Weinert & R. H. Kluwe (Eds.), *Metacognition, motivation, and understanding* (pp. 289-313). Hillsdale, NJ: Lawrence Erlbaum Associates.

- Gullberg, M., & Holmqvist, K. (2006). What speakers do and what addressees look at: Visual attention to gestures in human interaction live and on video. *Pragmatics and Cognition*, 14, 53-82.
- Gullberg, M., & Kita, S. (2009). Attention to speech-accompanying gestures: Eye movements and information uptake. *Journal of Nonverbal Behavior*, 33, 251-277.
- Gurney, D. J., Pine, K. J., & Wiseman, R. (2013). The gestural misinformation effect: skewing eyewitness testimony through gesture. *The American Journal of Psychology*, 126, 301-314.
- Hazzan, O. (1999). Reducing abstraction level when learning abstract algebra concepts. *Educational Studies in Mathematics*, 40, 71-90.
- Henke, K. (2010). A model for memory systems based on processing modes rather than consciousness. *Nature Reviews Neuroscience*, 11, 523-532.
- Hostetter, A. B. (2011). When do gestures communicate? A meta-analysis. *Psychological Bulletin*, 137, 297.
- Iverson, J. M., & Goldin-Meadow, S. (1998). Why people gesture when they speak. *Nature*, 396, 228.
- Kaminski, J. A., Sloutsky, V. M., & Heckler, A. F. (2008). The advantage of abstract examples in learning math. *SCIENCE*, 320, 454-455.
- Keane, M., Gabrieli, J., Monti, L., Fleischman, D., Cantor, J., & Noland, J. (1997). Intact and impaired conceptual memory processes in amnesia. *Neuropsychology*, 11, 59-69.
- Kelly, S. D., Kravitz, C., & Hopkins, M. (2004). Neural correlates of bimodal speech and gesture comprehension. *Brain and language*, 89, 253-260.
- Kelly, S. D., Özyürek, A., & Maris, E. (2010). Two sides of the same coin: Speech and gesture manually interact to enhance comprehension. *Psychological Science*, 21, 260-267.
- Kelly S. D., McDevitt T., & Esch M. (2009). Brief training with co-speech gesture lends a hand to word learning in a foreign language. *Language and Cognitive Processes*, 24, 313-334.
- Kelly, S. D., Barr, D. J., Church, R. B., & Lynch, K. (1999). Offering a hand to pragmatic understanding: The role of speech and gesture in comprehension and memory. *Journal of memory and Language*, 40, 577-592.
- Kendon, A. (1980). Gesticulation and speech: Two aspects of the process of utterance. *The relationship of verbal and nonverbal communication*, 25, 207-227.

- Klooster, N. B., Cook, S. W., Uc, E. Y. & Duff, M. C. (2014). Gestures make memories, but what kind? Evidence of dissociations in amnesic and Parkinson's patients. Paper presentation at the International Society for Gesture Studies Conference, San Diego.
- Klooster, N. B., Cook, S. W., Uc, E. Y. & Duff, M. C. (2015). Gestures make memories, but what kind? Patients with impaired procedural memory display disruptions in gesture production and comprehension. *Frontiers in Human Neuroscience*, 8, 1054.
- Knowlton, B. J., & Squire, L. R. (1993). The learning of categories: Parallel brain systems for item memory and category knowledge. *Science*, 262, 1747-1749.
- Knowlton, B. J., & Squire, L. R. (1994). The information acquired during artificial grammar learning. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, 20, 79.
- Knowlton, B. J., & Squire, L. R. (1996). Artificial grammar learning depends on implicit acquisition of both abstract and exemplar-specific information. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, 22, 169.
- Knowlton, B., Squire, L., & Gluck, M. (1994) Probabilistic classification learning in amnesia. *Learning and Memory*, 1, 106-120.
- Lampinen, A. K., & McClelland, J. L. (2018). Different presentations of a mathematical concept can support learning in complementary ways. *Journal of Educational Psychology*, 110, 664-682.
- Lewicki, P., Hill, T., & Bizot, E. (1988). Acquisition of procedural knowledge about a pattern of stimuli that cannot be articulated. *Cognitive Psychology*, 20, 24-37.
- Loehr, D. (2007). Aspects of rhythm in gesture and speech. *Gesture*, 7, 179-214.
- Mathews, R. C., Buss, R. R., Stanley, W. B., Blanchard-Fields, F., Cho, J. R., & Druhan, B. (1989). Role of implicit and explicit processes in learning from examples: A synergistic effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 15, 1083-1100.
- McClelland, J. L., McNaughton, B. L., & O'Reilly, R. C. (1995). Why there are complementary learning systems in the hippocampus and neocortex: Insights from the successes and failures of connectionist models of learning and memory. *Psychological Review*, 102, 419.
- McNeil, N., Alibali, M., & Evans, J. L. (2000). The role of gesture in children's language comprehension: Now they need it, now they don't. *Journal of Nonverbal Behavior*, 24, 131-150.
- McNeill, D. (1992). *Hand and mind: What gestures reveal about thought*. Chicago, IL: University of Chicago Press.

- Merikle, P. M., & Reingold, E. M. (1991). Comparing direct (explicit) and indirect (implicit) measures to study unconscious memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 17, 224-233.
- Millward, R. B., & Reber, A. S. (1972). Probability learning: Contingent-event schedules with lags. *American Journal of Psychology*, 85, 81-98.
- Millward, R. B., & Reber, A. S. (1968). Event-recall in probability learning. *Journal of Verbal Learning and Verbal Behavior*, 7, 980-989.
- Milner, B., Corkin, S., & Teuber, H. L. (1968). Further analysis of the hippocampal amnesia syndrome: 14- year follow-up study of H.M. *Neuropsychologia*, 6, 215-235.
- Morgan, J. L., Meier, R. P., & Newport, E. L. (1987). Structural packaging in the input to language learning: Contributions of prosodic and morphological marking of phrases to the acquisition of language. *Cognitive Psychology*, 19, 498-550.
- Morrel-Samuels, P., & Krauss, R. M. (1992). Word familiarity predicts temporal asynchrony of hand gestures and speech. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 18, 615-622.
- Moss, A. J., Rosenzweig, C., Robinson, J., Litman, L. (2020). Demographic Stability on Mechanical Turk Despite COVID-19. *Trends in Cognitive Science*, 24, 678-680.
- Mumford, K. H., & Kita, S. (2014). Children use gesture to interpret novel verb meanings. *Child Development*, 85, 1181-1189.
- Musen, G. & Treisman, A. (1990). Implicit and Explicit Memory for Visual Patterns. *Journal of Experimental Psychology: Learning Memory and Cognition*, 16, 127-137.
- Nissen, M. J., & Bullemer, P. (1987). Attentional requirement of learning: evidence from performance measures. *Cognitive Psychology*, 19, 1-32.
- Nobe, S. (2000). Where do *most* spontaneous representational gestures actually occur with respect to speech? In D. McNeill (ed.), *Language and gesture* (pp. 186-198). N.Y.: Cambridge University Press.
- Novack, M. A., Congdon, E., Hemani-Lopez, N., & Goldin-Meadow, S. (2014). From action to abstraction: using the hands to learn math. *Psychological Science*, 25, 903-910.
- Olson, M. A., & Fazio, R. H. (2006). Reducing automatically activated racial prejudice through implicit evaluative conditioning. *Personality and Social Psychology Bulletin*, 32, 421-433.
- Perry, M. (1991). Learning and transfer: Instructional conditions and conceptual change. *Cognitive Development*, 6, 449-468.

- Perry, M., Berch, D., & Singleton, J. L. (1995). Constructing shared understanding: The role of nonverbal input in learning contexts. *Journal of Contemporary Legal Issues*, 6, 213-236.
- Ping, R., & Goldin-Meadow, S. (2008). Hands in the air: Using ungrounded iconic gestures to teach children conservation of quantity. *Developmental Psychology*, 44, 1277.
- Ping, R., Goldin-Meadow, S., & Beilock, S. (2014). Understanding gesture: Is the listener's motor system involved? *Journal of Experimental Psychology: General*, 143, 195-204.
- Plant, E. A., Peruche, B. M., & Butz, D. A. (2005). Eliminating automatic racial bias: Making race on-diagnostic for responses to criminal suspects. *Journal of Experimental Social Psychology*, 41, 141-156.
- Prather, R. W. (2012). Implicit learning of arithmetic regularities is facilitated by proximal contrast. *PLoS One*, 7, e48868.
- Rader, N. de V., & Zukow-Goldring, P. (2010). How the hands control attention during early word learning. *Gesture*, 10, 202-221.
- Rau, M. A. (2016). Conditions for the effectiveness of multiple visual representations in enhancing STEM learning. *Educational Psychology Review*, 29, 717-761.
- Reber, P. J. (2013). The neural basis of implicit learning and memory: A review of neuropsychological and neuroimaging research. *Neuropsychologia*, 51, 2026-2042.
- Reber, A. S. (1993). *Implicit learning and tacit knowledge: An essay on the cognitive unconscious*. New York: Oxford.
- Reber, A. S. (1989). Implicit learning and tacit knowledge. *Journal of Experimental Psychology: General*, 118, 219-235.
- Reber, A. S. (1976). Implicit learning of artificial languages: The role of instructional set. *Journal of Experimental Psychology: Human Learning and Memory*, 2, 88-94.
- Reber, A. S. (1967). Implicit learning of artificial grammars. *Journal of Verbal Learning and Verbal Behavior*, 6, 855-863.
- Reber, P.J., Knowlton, B. J., & Squire, L. R. (1996) Dissociable properties of memory systems: Differences in the flexibility of declarative and nondeclarative knowledge. *Behavioral Neuroscience*, 110, 859-869.
- Reber, A., & Lewis, S. (1977). Toward a theory of implicit learning: The analysis of the form and structure of a body of tacit knowledge. *Cognition*, 5, 333-361.
- Reber, A. S., & Millward, R. B. (1971). *Event tracking in probability learning*. *American Journal of Psychology*, 84, 85-99.

- Reber, A. S., & Millward, R. B. (1968). Event observation in probability learning. *Journal of Experimental Psychology*, 77, 317-327.
- Reber, A. S., & Millward, R. B. (1965). *Probability learning and memory for event sequences*. *Psychonomic Science*, 3, 431-432.
- Reber, P. J., & Squire, L. J. (1994). Parallel brain systems for learning with and without awareness. *Learning and Memory*, 1, 217-229.
- Reber, P. J., & Squire, L. J. (1998). Encapsulation of implicit and explicit memory in sequence learning. *Journal of Cognitive Neuroscience*, 10, 248-263.
- Reber, A. S., Kassin, S. M., Lewis, S., & Cantor, G. (1980). On the relationship between implicit and explicit modes of learning a complex rule structure. *Journal of Experimental Psychology: Human Learning and Memory*, 6, 492-502.
- Reed, J., & Johnson, P. (1994). Assessing implicit learning with indirect tests: determining what is learned about sequence structure. *Journal of Experimental Psychology*, 20, 585-594.
- Richland, L. E., Stigler, J. W., & Holyoak, K. J. (2012). Teaching the conceptual structure of mathematics. *Educational Psychologist*, 47, 189-203.
- Rittle-Johnson, B., Schneider, M., & Star, J. R. (2015). Not a one-way street: bidirectional relations between procedural and conceptual knowledge of mathematics. *Educational Psychology Review*, 27, 587-597.
- Roediger, H. L., & McDermott, K. B. (1993). Implicit memory in normal human subjects. In F. Boller & J. Grafman (Eds.), *Handbook of Neuropsychology* (Vol. 8, pp. 63-131). Amsterdam: Elsevier.
- Rydell, R. J., McConnell, A. R., & Mackie, D. M. (2008). Consequences of discrepant explicit and implicit attitudes: Cognitive dissonance and increased information processing. *Journal of Experimental Social Psychology*, 44, 1526-1532.
- Schneider, M., & Stern, E. (2010). The developmental relations between conceptual and procedural knowledge: a multimethod approach. *Developmental Psychology*, 46, 178-192.
- Schwartz, D. L., & Goldstone, R. L. (2015). Learning as coordination. In L. Corno & E. M. Anderman (Eds.), *Handbook of Educational Psychology* (3rd Ed., Chap. 5, pp. 61-75). New York, NY: Routledge.
- Schwartz, N. D. (2020, May 8). How bad is unemployment? ‘Literally off the charts’. *The New York Times*; 2020. nytimes.com.

- Scoville, W. B., and Milner, B. (1957). Loss of recent memory after bilateral hippocampal lesions. *Journal of Neurology, Neurosurgery & Psychiatry*, 20, 11-21.
- Seger, C. A. (1994). Implicit Learning. *Psychology Bulletin*, 115, 163-196.
- Shanks, D. R. (2005). *Implicit learning*. In K. Lamberts & R. Goldstone (Eds.), *Handbook of Cognition* (pp. 202-220). London, England: Sage.
- Shanks, D. R., Johnstone, T., & Staggs, L. (1997). Abstraction processes in artificial grammar learning. *Quarterly Journal of Experimental Psychology*, 50, 216-252.
- Shanks, D. R., & St. John, M. F. (1994). Characteristics of dissociable human learning systems. *Behavioral and Brain Sciences*, 17, 367-447.
- Shimpi, P. M., & Huttenlocher, J. (2007). Redirective labels and early vocabulary development. *Journal of Child Language*, 34, 845-859.
- Shintel, H., Nusbaum, H., & Okrent, A. (2006). Analog acoustic expression in speech communication. *Journal of Memory and Language*, 55, 167-177.
- Siegler, R. S. (1994). Cognitive variability: A key to understanding cognitive development. *Current Directions in Psychological Science*, 3, 1-5.
- Siegler, R., & Stern, E. (1998). Conscious and unconscious strategy discovery: a microgenetic analysis. *Journal of Experimental Psychology: General*, 127, 377-397.
- Simonite, T. (2020, April 3). Newly Unemployed, and Labeling Photos for Pennies. wired.com.
- Singer, M. A., & Goldin-Meadow, S. (2005). Children learn when their teachers' gestures and speech differ. *Psychological Science*, 16, 85-89.
- Squire, L. R. (2004). Memory systems of the brain: a brief history and current perspective. *Neurobiology of Learning and Memory*, 82, 171-177.
- Squire, L. R., & Zola, S. M. (1996). Structure and function of declarative and nondeclarative memory systems. *Proceedings of the National Academy of Sciences of the United States of America*, 93, 13515-13522.
- Stadler, M.A. (1995). Role of attention in implicit learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 21, 674-685.
- Stanley, W. B., Mathews, R. C., Buss, R. R., & Koller-Kope, S. (1989). Insight without awareness: On the interaction of verbalization, instruction, and practice in a simulated process control task. *Quarterly Journal of Experimental Psychology*, 41, 553-577.

- Sun, R., Merrill, E., & Peterson, T. (1998). A bottom-up model of skill learning. In *Proceedings of the 20th Cognitive Science Society Conference* (pp. 1037-1042). Mahwah, NJ: Lawrence Erlbaum Associates.
- Sun, R., Merrill, E., & Peterson, T. (2001). From implicit skill to explicit knowledge: A bottom-up model of skill learning. *Cognitive Science*, 25, 203-244.
- Sun, R., Slusarz, P., & Terry, C. (2005). The interaction of the explicit and the implicit in skill learning: A dual-process approach. *Psychological Review*, 112, 159-192.
- Thompson, L. A. (1995). Encoding and memory for visible speech and gestures: A comparison between young and older adults. *Psychology and Aging*, 10, 215.
- Valenzeno, L., Alibali, M. W., & Klatzky, R. L. (2003). Teachers' gestures facilitate students' learning: A lesson in symmetry. *Contemporary Educational Psychology*, 28, 187-204.
- Wakefield, E. M., Foley, A. E., Ping, R., Villarreal, J. N., Goldin-Meadow, S., & Levine, S. C. (2019). Breaking Down Gesture and Action in Mental Rotation: Understanding the Components of Movement That Promote Learning. *Developmental Psychology*, 55, 981-993.
- Wakefield, E. M., Hall, C., James, J., Goldin-Meadow, S. (2018a). Gesture for generalization: Gesture facilitates flexible learning of words for actions on objects. *Developmental Science*, 21, 1-14.
- Wakefield, E., Novack, M. A., Congdon, E. L., Franconeri, S., & Goldin-Meadow, S. (2018b). Gesture helps learners learn, but not merely by guiding their visual attention. *Developmental Science*, 21, e12664.
- Whittlesea, B. W. A., & Dorken, M. D. (1993). Incidentally, things in general are particularly determined: an episodic-processing account of implicit learning. *Journal of Experimental Psychology General*, 122, 227-248.
- Willingham, D., Nissen, M., & Bullemer, P. (1989). On the development of procedural knowledge. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 15, 1047-1060.
- Ziegler, E., Edelsbrunner, P. A. & Stern, E. (2018). The relative merits of explicit and implicit learning of contrasted algebra principles. *Educational Psychology Review*, 30, 531-558.

APPENDICES

Appendix A. Study 1 Training and Posttest Instructions

Training Instructions by Condition

1. Implicit Alone and Implicit+Gesture

Instructions given prior to training block 1:

In today's part of the experiment, you will watch many short videos. Each video will show an equation. You should look at the equation and try to remember it.

After the 2-minute break, prior to beginning training block 3:

Here is a reminder of the instructions for this part of the experiment:

You will now watch some more videos. Each video will show an equation. You should look at the equation and try to remember it.

2. Explicit Alone and Explicit+Gesture

Instructions given prior to training block 1:

In today's part of the experiment, you will watch many short videos. Each video will show an equation. You should look at the equation and try to remember it.

The equations you will see follow a set of rules. These rules allow certain symbols to combine to make other symbols. Your task is to try to figure out what these rules are—that is, which symbols combine to make other symbols. You will be tested on your knowledge of these rules, so it is important that you try your best to learn them.

After the 2-minute break, prior to beginning training block 3:

Here is a reminder of the instructions for this part of the experiment:

You will now watch some more videos. Each video will show an equation. You should look at the equation and try to remember it.

The equations you will see follow a set of rules. These rules allow certain symbols to combine to make other symbols. Your task is to try to figure out what these rules are—that is, which symbols combine to make other symbols. You will be tested on your knowledge of these rules, so it is important that you try your best to learn them.

3. Implicit→Explicit Alone and Implicit→Explicit+Gesture

Instructions given prior to training block 1:

In today's part of the experiment, you will watch many short videos. Each video will show an equation. You should look at the equation and try to remember it.

After the 2-minute break, prior to beginning training block 3:

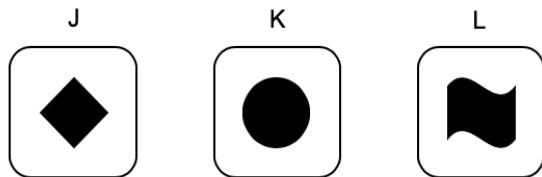
Please read the NEW instructions below for the next part of the experiment:

You will now watch some more videos. Each video will show an equation. You should look at the equation and try to remember it.

The equations you will see follow a set of rules. The rules allow certain symbols to combine to make other symbols. Your task is to try to figure out what these rules are—that is, which symbols combine to make other symbols. You will be tested on your knowledge of these rules, so it is important that you try your best to learn them.

Figure 33. Participants in all conditions also received the following instructions prior to beginning training block 1.

After each video plays, a new page will automatically load (this may take a few seconds). On the new page, you will see the same equation from the video, but with a symbol missing. Below each problem, the symbols will be shown with their corresponding key, like this:



Place your right hand on the "J," "K," and "L" keys, like this:



You should press the key that matches the symbol that correctly solves the equation you saw in the video. After you make your choice, the correct answer will be given. Halfway through watching the videos, you will have the option to take a 2-minute break.

Posttest Instructions

Figure 34. Participants in all conditions received the following instructions prior to beginning the posttest.

You will now see some more equations. These equations follow the same set of rules as the equations from the videos you just watched.

You will be shown 3 equations at a time and asked to choose which one of the equations is solved correctly. The equations will be labeled "J," "K," and "L," like this:

Which equation is solved correctly?

$$\bullet, \diamond, \text{wavy} \rightarrow \bullet, \diamond$$

J

$$\text{wavy}, \text{wavy}, \bullet \rightarrow \diamond, \text{wavy}$$

K

$$\bullet, \bullet, \bullet \rightarrow \text{wavy}, \bullet$$

L

You should press the key that matches the equation you think is solved correctly. You may submit your answer for each question after 3 seconds have passed, but take as much time as you need.

Some of the questions will be more challenging than others. Just try your best.

Appendix B. Study 1 Questionnaire

We would now like to ask you a few questions about what you thought of the experiment.

- 1. What do you think was the goal of this experiment?*
- 2. After you watched the videos yesterday, you did a task in which you had to choose which equations were solved correctly. Did you notice anything about the videos that led you to decide whether or not each equation was solved correctly?*
- 3. If the speaker used gestures in the videos you watched yesterday, what did you think of them? Did you find them helpful? Confusing? Or did you not really notice them?*
- 4. If you were told to learn the rules for solving the equations at the beginning of the experiment yesterday, or you think you may have an idea of what the rules are, please list them here. It's okay if you are uncertain about your answers.*

Appendix C. Study 2 and 3 Rule Video Images and Scripts

Figure 35. Video 1 – “Rule 1: The order of the symbols on the left does not change the result. For example, triangle combined with square makes triangle, and square combined with triangle also makes triangle.”

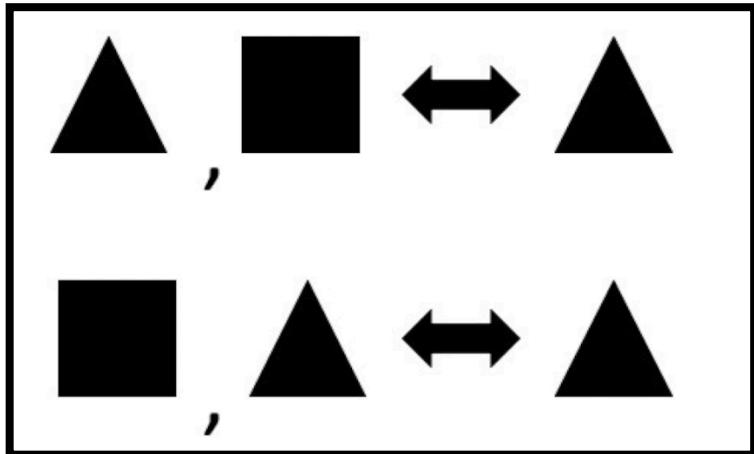


Figure 36. Video 2 – “Rule 2: When any symbol combines with square, the result will always be *that* symbol. For example, square combined with triangle makes triangle, and circle combined with square makes circle.”

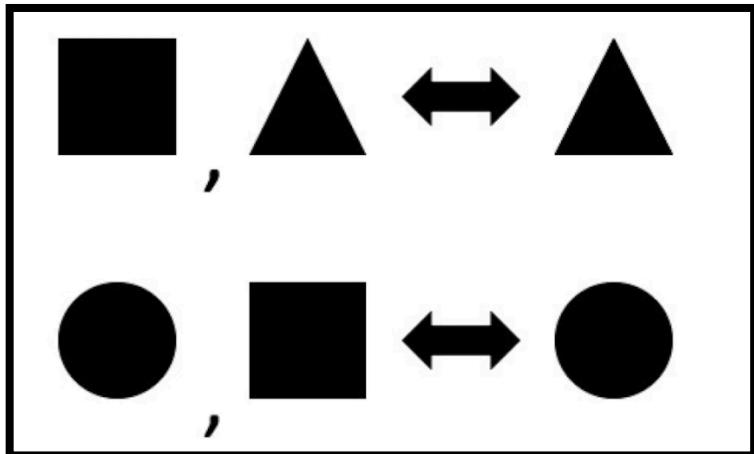


Figure 37. Video 3 – “Rule 3: Circle combined with triangle makes square.”

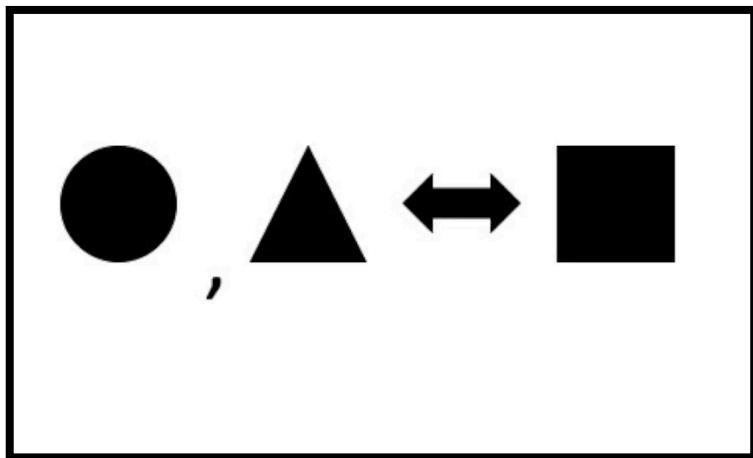


Figure 38. Video 4 – “Rule 4: Circle combined with circle makes triangle.”

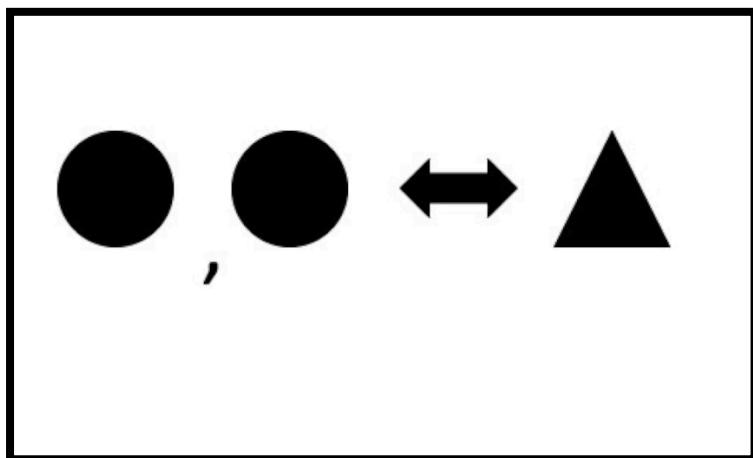


Figure 39. Video 5 – “Rule 5: Triangle combined with triangle makes circle.”

