COGNITIVE SCIENCE
A Multidisciplinary Journal

Cognitive Science 44 (2020) e12889

ISSN: 1551-6709 online
DOI: 10.1111/cogs.12889

Learning From Gesture and Action: An Investigation of Memory for Where Objects Went and How They Got There

Autumn B. Hostetter,[a] Wim Pouw,[b,c] Elizabeth M. Wakefield[d]

[a]*Department of Psychology, Kalamazoo College*
[b]*Donders Institute for Brain, Cognition and Behaviour, Radboud University Nijmegen*
[c]*Max Planck Institute for Psycholinguistics, Nijmegen*
[d]*Department of Psychology, Loyola University Chicago*

Received 18 March 2019; received in revised form 10 July 2020; accepted 27 July 2020

Abstract

Speakers often use gesture to demonstrate how to perform actions—for example, they might show how to open the top of a jar by making a twisting motion above the jar. Yet it is unclear whether listeners learn as much from seeing such gestures as they learn from seeing actions that physically change the position of objects (i.e., actually opening the jar). Here, we examined participants' implicit and explicit understanding about a series of movements that demonstrated how to move a set of objects. The movements were either shown with actions that physically relocated each object or with gestures that represented the relocation without touching the objects. Further, the end location that was indicated for each object covaried with whether the object was grasped with one or two hands. We found that memory for the end location of each object was better after seeing the physical relocation of the objects, that is, after seeing *action*, than after seeing gesture, regardless of whether speech was absent (Experiment 1) or present (Experiment 2). However, gesture and action built similar implicit understanding of how a particular handgrasp corresponded with a particular end location. Although gestures miss the benefit of showing the end state of objects that have been acted upon, the data show that gestures are as good as action in building knowledge of *how* to perform an action.

Keywords: Gesture; Action; Comprehension

Correspondence should be sent to Autumn B. Hostetter, Department of Psychology, Kalamazoo College, 1200 Academy Street, Kalamazoo, MI 49006. E-mail: autumn.hostetter@kzoo.edu

1. Introduction

Speakers often accompany their spoken message with gestures: hand movements that represent information about actions, objects, or events. These hand movements serve an important purpose for listeners—When speakers gesture, listeners have better comprehension of their message than when speakers do not gesture. For example, in a meta-analysis of 63 studies that compared the effectiveness of speech presented alone versus speech presented with gesture, Hostetter (2011) found a significant moderate benefit of gesture (see also Dargue, Sweller, & Jones, 2019). Although speakers use gesture ubiquitously in communication (e.g., McNeill, 1992), it is not the only type of communicative movement speakers engage in. Speakers may also act on objects in their physical environment to convey information about those objects, yet few studies have directly compared the effects of observing gestures about objects with the effects of observing actions on objects (though see Kelly, Healey, Özyürek, & Holler, 2015). In the present study, we examine the mnemonic effects of seeing gestures that show how and where to move objects to the effects of seeing actions that actually move those objects.

The information gleaned from gesture versus action may be different, because listeners interpret a different purpose for gesture than for action. Specifically, actions performed on objects are interpreted as goal-directed, with observers attending more to what happened as a result of a particular action than to how it happened. This preference is observed from the first year of life; if an infant observes an actor reach repeatedly for one object, they notice if the actor then changes her goal (reaching for a new object) but they do not notice if the actor changes the movement pattern to get to that goal (reaching for the same object in a new location; Buresh & Woodward, 2007; Woodward, 1998). This cognitive bias continues across development: When seeing an action occurring on an object, adults focus on what the action achieves, not the particular movement patterns of the action itself (Novack, Wakefield, & Goldin-Meadow, 2016; Schachner & Carey, 2013).

In contrast, gestures are schematic (Kita, Alibali, & Chu, 2017) and may preferentially highlight *particulars* of a movement pattern. Because gestures are not constrained by the same physical parameters as actions on objects, a speaker has more flexibility in terms of which aspects of a movement pattern to show in gesture than in action. For example, using action to demonstrate where to place a heavy object may necessitate using two hands to lift the object, while indicating the desired location in gesture could easily be accomplished with a single hand. As such, listeners may assume that a speaker who chooses to gesture with two hands is doing so intentionally, to show *how* the object should be lifted and not just where it should go. In other words, the two-handedness of the gesture is interpreted as communicating something about the way the object is to be moved, while the two-handedness of the action is not interpreted as an important feature of meeting the end goal. In line with this possibility, there is some evidence that adults systematically interpret gestures as having a communicative intent to show *how* an action should be done. Specifically, Novack et al. (2016) found that observers were more likely to infer a representational intention when an actress produced a gesture that was near objects than when she produced an action on objects. The idea that gestures are recruited

to selectively highlight *how* a movement is performed rather than its outcome aligns with accounts suggesting that gestures emerged from early-hominin tool-making apprenticeships (e.g., teaching stone-knapping), wherein gestures could demonstrate how to complete an action, without having to physically hold the object or tool (Cataldo, Migliano, & Vinicius, 2018; Gärdenfors, 2017).

Additional evidence that gestures may highlight the particulars of a movement, whereas actions highlight the overarching goal of movement itself, can be found in the developmental literature. Mumford and Kita (2014) found that whether children were taught a novel verb through action or through gesture affected how they later generalized the verb in a new situation. When children had only seen the action, they most frequently generalized the verb to another action that would produce the same end state. Thus, the particular movement pattern was not encoded as being important; rather, children saw the movement in terms of the intended goal—the end state. In contrast, when they were taught the novel verb with a gesture that highlighted the manner in which the action was performed, the children were more likely to generalize the verb to another action that had a similar manner (rather than a similar end state). This shows that gesture can direct children's attention to how the movement was carried out, a feature that would encourage a focus on the particular features of the movement as important to the event. This same phenomenon was demonstrated when the actions being schematized by gesture were full-body movements performed by an actor: Aussems and Kita (2017) showed children videos of actors moving in a particular manner (e.g., skipping, trotting, hopping), and showed some children gestures that schematized these movement patterns. Seeing gesture as opposed to whole-body enactments boosted children's memory for how the actor moved. Further, Wakefield, Hall, James, and Goldin-Meadow (2018) found that seeing (or producing) gesture led children to more adeptly generalize the meaning of a novel verb than seeing (or producing) action. Thus, it appears that gesture may be uniquely situated to highlight information about manner of movement or particular movement patterns, whereas action emphasizes end state and goal-oriented parts of a movement event.

If gestures do highlight movement patterns more effectively than action, a further question is whether they do so through implicit or explicit processes. Implicit processes are those that occur largely without conscious awareness and result in knowledge that cannot be overtly stated (Dienes & Berry, 1997; Reber, 1989). For example, procedural memory for a motor routine (i.e., walking, playing a piece on the piano, riding a bike) may be difficult to describe through language, even while the motor task can be performed without error. In contrast, explicit processes occur when learning is above participants' subjective threshold (i.e., they know that they know and can express this through language; see Dienes & Berry, 1997). The majority of research investigating the effects of gesture on learning has measured explicit knowledge, for example, by asking participants to state what they know (e.g., Cook, Duffy, & Fenn, 2013; van Wermeskerken, Fijan, Eielts, & Pouw, 2016) or to solve problems after receiving instruction incorporating gesture and explain how they arrived at their answers (e.g., Wakefield, Novack, Congdon, Franconeri, & Goldin-Meadow, 2018). Such research suggests that the effects of gesture are often apparent on tasks that require explicit use of the information.

On the other hand, some research suggests that gestures may primarily affect learning through implicit (rather than explicit) processes. For example, Alibali, Flevares, and Goldin-Meadow (1997) provide evidence that gesture can change behavior without an individual's awareness. In the study, adults were told to assess a child's knowledge about a math concept by talking with the child about their incorrect solution to a math equivalence problem. The adult then provided instruction to the child about how to correctly solve the problem. Alibali et al. (1997) found that adults adjusted the number of strategies they taught a child, based on whether the child expressed some understanding of how to solve the problem in their gesture. However, most adults said that they were unaware that children were gesturing during the study; thus, the children's gestures likely affected adults' behavior through implicit means. Children's behavior is also affected by the gestures of other children (e.g., Kelly & Church, 1997) and of adults (Singer & Goldin-Meadow, 2005), with children processing information that is presented in gesture and not in speech. It remains unclear whether children's awareness of the information conveyed in gesture is explicit or implicit.

Perhaps the clearest evidence that gesture affects learning and behavior through implicit processes comes from work with clinical populations who have damage to one memory system but not the other. For example, patients diagnosed with hippocampal amnesia have intact implicit procedural memory but impaired declarative memory. Hilverman, Cook, and Duff (2018) taught patients with hippocampal amnesia and healthy controls novel label-object pairings with or without the support of gesture. Although healthy controls could learn words under all conditions, patients with hippocampal amnesia could only remember words if they had produced gestures while encoding the pairs. This suggests that gestures may engage implicit memory systems (the system still intact for the patients with hippocampal amnesia), at least when they are produced by the learner. Complementing these findings, Klooster, Cook, Uc, and Duff (2014) found that patients with the *opposite* pattern of memory impairment—those with Parkinson's disease who have impaired implicit/procedural memory but intact explicit/declarative memory—were not impacted by gesture they had *perceived*. Patients with Parkinson's disease and healthy controls viewed explanations of how to solve a puzzle task, in which the speaker's gestures were either flat-sideways gestures or high-arched gestures and then solved the same puzzle task. Healthy controls used movement patterns that reflected the gestures they had seen, but this effect was absent in patients with Parkinson's disease. Again, this suggests that gestures might primarily affect implicit (rather than explicit) memory, as gestures had no effect for the patients with impaired implicit memory. However, more empirical attention should be given to this question, as it is unclear whether gestures also primarily affect implicit memory in healthy populations.

1.1. The present study

The present study had two primary goals. First, we aimed to test the hypothesis that information about how to perform an action will be highlighted more by gestures than by

actions on objects. Second, we aimed to explore whether the proposed benefit of gesture would be more evident on a measure of implicit or explicit learning.

Toward these goals, we developed a novel paradigm in which participants watched a series of videos showing a woman either physically moving (action condition) or representing how to move (gesture condition) objects to one of two positions (a top shelf or a bottom shelf). Participants were told that their memory for where the objects went would be tested. However, the videos that each participant saw actually followed a particular hand/end-location rule. Specifically, the videos either showed objects grasped with one hand being *moved up* and objects grasped with two hands being *moved down*, or vice versa (i.e., one hand moved down, two hands moved up). In this way, participants were exposed to gestures or actions that embodied a particular movement pattern between handedness and end location, without their attention being purposefully directed toward that aspect of the movement in either condition. If gestures are more effective than actions at highlighting movement patterns rather than end goals, we predict that participants will attune more to the correspondence between hand and end location after seeing gesture than after seeing action. We tested participants' understanding of the hand/end-location rule in several ways.

First, to test *implicit* understanding, we built on the finding that implicit, procedural memory is context-sensitive (e.g., Borghi & Riggio, 2015). We reasoned that participants who have good implicit understanding of the hand/end-location rule should have better memory for the correct end location of a particular object when the object is presented with the same handgrasp that was shown during training than when it is presented with the alternative handgrasp. That is, when the context (e.g., handgrasp) of the object shown at test is the same as the context shown during learning, memory for the end location should be better than when the context is different. To test this, we showed participants two still images of the woman holding each object—one depicted a congruent grasp (i.e., the same grasp that had been shown during training) and one an incongruent grasp (i.e., the alternative grasp from what had been shown during training). Participants were tasked with indicating where each object had been placed (top or bottom). We predict that participants will demonstrate better memory for the end location of each object when it is shown in the same context (i.e., with the same congruent handgrasp) that it was shown with during training than when it is shown with the alternative handgrasp. Further, if viewing gesture (vs. action) has made the implicit understanding of the hand/end-location rule stronger, then this congruency effect should be more pronounced in the gesture condition than in the action condition.

Participants' *explicit* understanding of the rule was tested with varying levels of scaffolding. If seeing gesture (rather than action) has made the correspondence between hand grasp and ending location more explicit for participants, those who saw gesture should have more success explicitly stating the rule than those who saw action. However, it is quite possible that participants' implicit understanding of the rule (as indexed by the congruency effect described above) could be affected by gesture, without it affecting their explicit understanding.

2. Experiment 1

2.1. Method

2.1.1. Participants

We used Amazon's mechanical turk (mturk) to target 200 participants currently residing in the United States. Participants were compensated $1.45 for completing the 10-min study. To arrive at our sample size, we took the amount of money we had available for the study and divided by the per person cost (equivalent of minimum wage plus the mturk fee). A priori (preregistered) power analysis (G*Power version 3.1) suggested that a sample size of 200 would result in 82% power to detect a small to medium effect size (w = 0.2) at an overall alpha of 0.05, and 62% power to detect an effect of that magnitude on any individual test (using an alpha level of 0.01 to correct for multiple measures).

Experiment 1 was completed by 201 people (107 men; 94 women). However, as planned in the preregistration (https://osf.io/je2sn/), data from participants who reported experiencing any issues with the videos or images were discarded (n = 5), as were data from participants who reported writing down the names of the objects or their locations during the training (n = 2). We also excluded data from one participant who reported not understanding the instructions and from 11 participants who did not answer the majority of the memory and transfer trials. Thus, the analyzed sample consisted of data from 182 participants (92 men; 90 women). Their average age was 36.97 years (SD = 11.81). The majority (77%) self-reported their ethnicity as White/Caucasian, with other participants identifying as Asian (7%), Black/African American (8%), Hispanic/Latinx (3%), or other (5%). All participants in the analyzed sample rated their proficiency with English as a 4 (n = 3) or 5 (n = 179) on a 5-point scale with 5 = fluent, and 86% reported being exposed to English from birth. Participants were randomly assigned to the action (n = 93) or gesture (n = 89) condition.

2.1.2. Stimuli

2.1.2.1. Objects: We gathered 16 objects for use in the training phase of the experiment. In choosing objects, we aimed to have an equal number of human-made versus natural objects that varied in size, but that could be lifted plausibly with either one or two hands. The 16 objects are listed in Table 1. We divided the objects into two sets, approximately matching the size and shape of each object in Set A with an object of similar size and shape in Set B. We also balanced the number of human-made and natural objects across the two sets. For example, sets A and B each contained four human-made and four natural objects, and each contained some relatively small things (e.g., rock, crystal) and some relatively large things (e.g., guitar/ukulele, bat). The particular exemplars that we chose could all be lifted with either one or two hands.

2.1.2.2. Training videos: A woman dressed in black was filmed against a neutral background. She stood behind a table facing the camera. On the table centered in front of the

Table 1
Objects used in the training phase of the experiments

Set A	Set B
Mushroom	Carrot
Guitar	Bat
Seashell	Pinecone
Apple	Potato
Light bulb	Candle
Rock	Crystal
Horse	Stapler
Hat	Cup

Note. Each participant saw all objects within a set grasped with the same handgrasp (one vs. two hand) and moved to the same ending location (top vs. bottom). Participants saw one of four rules: (a) objects in Set A grasped with one hand and moved up, objects in Set B grasped with two hands and moved down; (b) objects in Set A grasped with two hands and moved up, objects in Set B grasped with one hand and moved down; (c) objects in Set A grasped with one hand and moved down, objects in Set B grasped with two hands and moved up; (d) objects in Set A grasped with two hands and moved down, objects in set B grasped with one hand and moved up. Participants saw either action (objects actually grasped and moved) or gesture (hands indicated how to grasp and move each object) for all objects in both sets.

woman, there was a wooden apparatus that consisted of two shelves, one positioned at table level, just below the woman's waist, and one positioned 61 cm above that, just below the woman's shoulders. To the woman's right, beside the shelf apparatus, there was a platform that stood 46 cm above the table.

Each object was filmed once in each of the 2 (hand grasp: one vs. two) × 2 (ending location: up vs. down) × 2 (gesture vs. action) conditions, resulting in eight videos for each of the 16 objects (128 videos total). At the beginning of each video, one of the 16 objects (e.g., rock, apple) was positioned on the platform to the speaker's right, and the woman stood facing forward with her hands by her sides (see Fig. 1). In the action videos, she reached for and grasped the object with either one or two hands (see top panel of Fig. 1). She brought the object to a central position in front of her body, level with the top of the platform, so it was visible in between the upper and lower shelves on the apparatus. Finally, she placed the object either up on the top shelf or down on the bottom shelf and returned her hands to the starting position by her sides. The one-hand grasps were always performed with the woman's right hand (the hand closest to the starting position of the object) and were produced as she deemed necessary to most naturally lift the object. For example, the guitar/ukulele was lifted with a whole-hand power grasp around its neck, whereas the hat was lifted by cupping the top of the hat. The mechanics of the two hand grasps also varied depending on the object. In the gesture videos, she followed the same sequence of movements, except that instead of actually grasping the object, she showed how to grasp and move the object in gesture by mimicking the grasp and movement pattern used in the action videos, but not directly interacting with the object (see bottom panel of Fig. 1). The woman's face was blurred in all videos and the videos were played without sound. Each of the 128 videos was between 4.6 and 7.0 s long (M = 5.36 s, SD = 0.49).

In the experiment, each participant saw one video for each of the 16 objects. All 16 videos seen by an individual participant were either action videos or gesture videos. Further, in the videos each participant saw, all eight objects from a particular set were moved with the same handgrasp to the same end location and all eight objects from the other set were moved with the opposite handgrasp to the opposite end location, though the particular handgrasp and end location assigned to each set of objects was counterbalanced across participants. To accomplish this, we created four presentation conditions that included all possible pairings of a particular object set, handgrasp, and end location (e.g., Set A moved with one hand to the top; Set A moved with one hand to the bottom). We then randomly assigned participants to view one of these four presentation conditions in either the gesture or the action condition. In this way, each individual object was shown being grasped with one versus two hands and being moved to the top versus bottom equally often across all participants. Further, because the objects in each set were mixed in terms of size, shape, and semantic category, the only rule to explain consistently the differentiation in whether the object was moved up or down was the handgrasp the woman used.[1] Within each presentation condition, the 16 videos were shown in a random order.

Fig. 1. Still images showing the progression of events in an Action video and the corresponding Gesture video. Four versions of each video were created, with each object being placed on the top and on the bottom with a one-hand and a two-hand grasp.

2.1.2.3. Still images: A still image was taken from each action video that showed the woman's hands and shoulders as she held the object in central position, before placing it on either the upper or lower shelf (see center top panel of Fig. 1). In addition to the images of the woman holding the 16 objects used in the training videos, the woman was also photographed holding 16 novel objects (e.g., flower, toy bear), for use in the transfer trials (see Table 1 of Appendix S1). The novel objects were held with the hand grasp (one vs. two hands) that they most naturally afforded, with eight objects being held with one hand and eight with two hands.

2.1.3. Procedure

The experiment was advertised as a study about memory. Interested participants were directed to a Qualtrics survey where they first certified that they were at least 18 years of age and that they were at a computer with a keyboard (i.e., not a tablet or mobile device). There were multiple steps to the procedure, as shown in Fig. 2 and described in detail below. Participants spent about 10 min (M = 615 s, SD = 215) completing the entire procedure.

2.1.3.1. Training phase: Participants were told that they would see videos of a woman and that their memory would be tested later. Participants were randomly assigned to view the 16 training videos in either the gesture or the action condition. The particular hand/

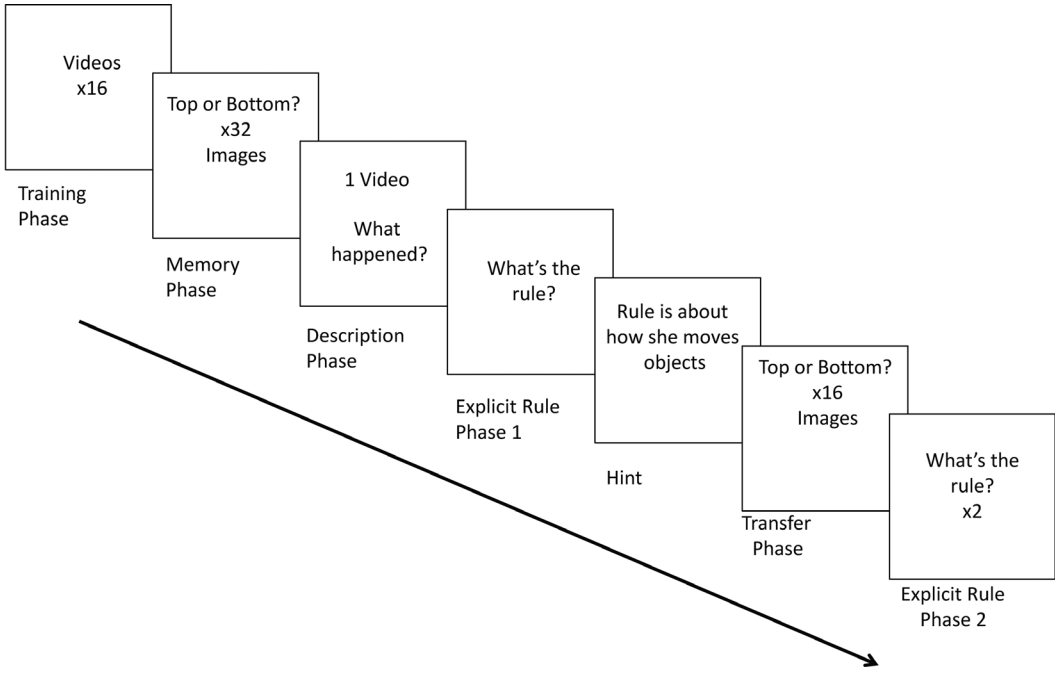

Fig. 2. The progression of events in the procedure of both experiments.

end-location rule demonstrated was counterbalanced across participants, as were the particular objects that were shown being moved in each way (see Section 2.1.2). The videos played in the center of the screen, one after the other, in a random order, and were presented at a size of 480 × 640 pixels.

2.1.3.2. Memory phase: Participants then completed the memory phase. Participants were told that they would see still images of the woman holding each object and they should recall as quickly[2] and accurately as possible whether the object had been associated with the top shelf or with the bottom shelf in the training videos. Participants indicated top by pressing T on the keyboard and indicated bottom by pressing B. The T and B keys were chosen because their vertical orientation on the keyboard aligns with the up and down direction of the woman's movement. Each of the 16 objects was shown twice for a total of 32 trials. In one image shown for each object (i.e., the congruent trials), the woman held the object with the same handgrasp as shown in the training phase. In the other image shown for the object (i.e., the incongruent trials), she held the object with the alternative handgrasp. The 32 images were presented in a random order at a size of approximately 200 × 140 pixels, and participants did not receive feedback about their responses.

2.1.3.3. Description phase: Participants were then shown one of the training videos again, to refamiliarize them with the structure of the training videos before moving on to the first explicit rule phase. They were asked to type a description of what happened in the video.

2.1.3.4. First explicit rule phase: Participants were then told that the woman actually had preferences for which objects should go up to the top shelf and which objects should go down to the bottom shelf. They were asked to type their single best guess describing the rule dictating the woman's preferences and press return.

2.1.3.5. Hint: Participants were told that the woman's rule had to do with how she moved the objects. During pilot testing, we found that most participants offered an inaccurate guess about the woman's rule during the first explicit attempt (e.g., color of the objects, the position of the object's first initial in the alphabet), and then continued to rely on this rule throughout the rest of the experiment. Because this prevents participants from potentially going on to generate the correct rule, we introduced a hint at this point in the procedure as a means of giving participants minimal feedback about their first explicit guess, in the hopes of discouraging them from using an inaccurate rule to complete the transfer task.

2.1.3.6. Transfer phase: Following the hint, participants were told that they would see pictures of the woman holding new objects that had not been seen previously and that they should indicate as quickly as possible whether each object should go up to the top shelf (by pressing T) or down to the bottom shelf (by pressing B). They were also told that it was okay if they did not know for sure what the woman's preference is. They should make their decision as quickly as possible based on what they saw before. The 16

transfer objects were then shown one at a time in a random order. Each was presented in the center of the screen at an approximate size of 200 × 140 pixels.

2.1.3.7. Second explicit rule phase: Participants were reminded that the woman's preferences regarding where to place each object were determined by a rule and that the rule had to do with how she moved the objects. They were asked to type their single best guess for what the woman's rule was regarding whether a particular object should be placed on the top or bottom shelf. They were told that this could be the same guess they had made previously, or something new if their best guess had changed. After their response was recorded, they were invited to describe any additional rule possibilities that came to mind. In this way, we hoped to encourage participants to keep thinking beyond any initial incorrect explicit guesses, to see if they could come to the rule about the correspondence between handgrasp and end location with some prompting and continued reflection.

2.1.3.8. Participant information: Finally, participants were asked to describe their age, gender, and ethnicity, and were asked to rate their proficiency with English on a 5-point scale with 5 = fluent. They were also asked to report whether they had experienced any problems with the videos or images in the study, whether they had used any techniques other than their memory to remember the objects or their locations (e.g., writing them down), whether they were using a QWERTY keyboard, and whether they had experienced any other issues that they felt should be disclosed (e.g., being interrupted in the middle). They were thanked for their participation and given instructions about how to receive their compensation in mturk.

2.1.4. Data coding and exclusion

2.1.4.1. Memory phase: Participants' responses to the memory trials were scored as correct or incorrect depending on the videos they saw during the training. That is, if they saw the woman place the apple on the top (or gesture about it going up), then T responses for the pictures of the apple were considered correct and B responses were considered incorrect. Further, each trial was coded as congruent or incongruent based on whether the image showed the woman holding the object with the same or different handgrasp that had been shown to the participant for that object during the training phase. These scoring procedures were automated using R code available at https://osf.io/vwhgq/.

As planned in the preregistration, we excluded individual trials on the memory task for which the reaction time was >5 s (indicating the participant was likely distracted or disengaged on that trial) or recorded by Qualtrics as 0 (indicating that the participant pressed a key before the image appeared on the screen). These criteria resulted in the exclusion of 8.1% of the data (473 of 5,824 trials). The average number of trials included per participant was 29.4 (out of 32 maximum).

2.1.4.2. Explicit rule phases 1 and 2: Participants were given three opportunities to describe the rule governing the woman's movements (once before the hint that the rule

has to do with how she moved the objects and twice after the hint). We distinguished between descriptions given in the first phase (before the hint) and in the second phase (after the hint). In both explicit rule phases, the primary coding decision was whether the rule description indicated handgrasp as a relevant feature.

For responses given in Explicit Phase 1, we took a liberal approach to coding the explicit responses because pilot data suggested that people are very unlikely to get the handgrasp/end-location rule without a hint. As planned in the preregistration (https://osf.io/je2sn/), we gave credit for any response that mentioned anything about a correspondence between how the object was grasped and the direction it was moved. This included stating the rule exactly (e.g., "two hand objects went up") but also included more vague responses (e.g., "how she grasped it"). Note that we also coded responses that mentioned the correspondence incorrectly (e.g., participant said "objects grasped with two hands went up" when they had actually seen objects grasped with two hands go down) as identifying the handgrasp rule, because it suggests that participants understand that handgrasp is the relevant feature.

Responses given during the second explicit phase (after the hint that the rule has to do with how she moved the objects) were scored with a more conservative approach. Now, only responses that specifically mentioned the correspondence between one- versus two-hand grasp and ending location were counted as identifying the handgrasp/end-location rule. However, we again counted responses that conveyed the incorrect correspondence (e.g., "two hand objects go up" when objects lifted with two hands had actually gone down) because such responses suggest that the participant does understand that handgrasp with one versus two hands was the relevant feature. The pattern of results reported does not change if such responses are not counted as conveying the handgrasp rule.

2.1.5. Data analysis

Our prediction is that participants who see gesture will be more likely to learn the sensorimotor regularity between direction and handgrasp than those who see action. We included a variety of measures that could tap into such an understanding and did not have any a priori hypotheses about which of these measures would be most likely to show a difference between seeing action versus seeing gesture. To control Type I error across so many analyses, we adopted a conservative alpha level of 0.01 for all analyses. We followed the procedures specified in the preregistration (see https://osf.io/je2sn/) to conduct both confirmatory and exploratory analyses (analysis code available at https://osf.io/de4rm/). However, for the sake of brevity, we present only the analyses of the data from the memory phase (the implicit measure) and the two explicit phases in this report. The other analyses are described in Appendix S1 and do not change the overall conclusions.

2.2. Results and discussion

2.2.1. Performance on the memory task

The proportion of trials answered correctly by each participant on the memory task is shown in Fig. 3. Chance performance in each condition is 0.50. Perfect memory for

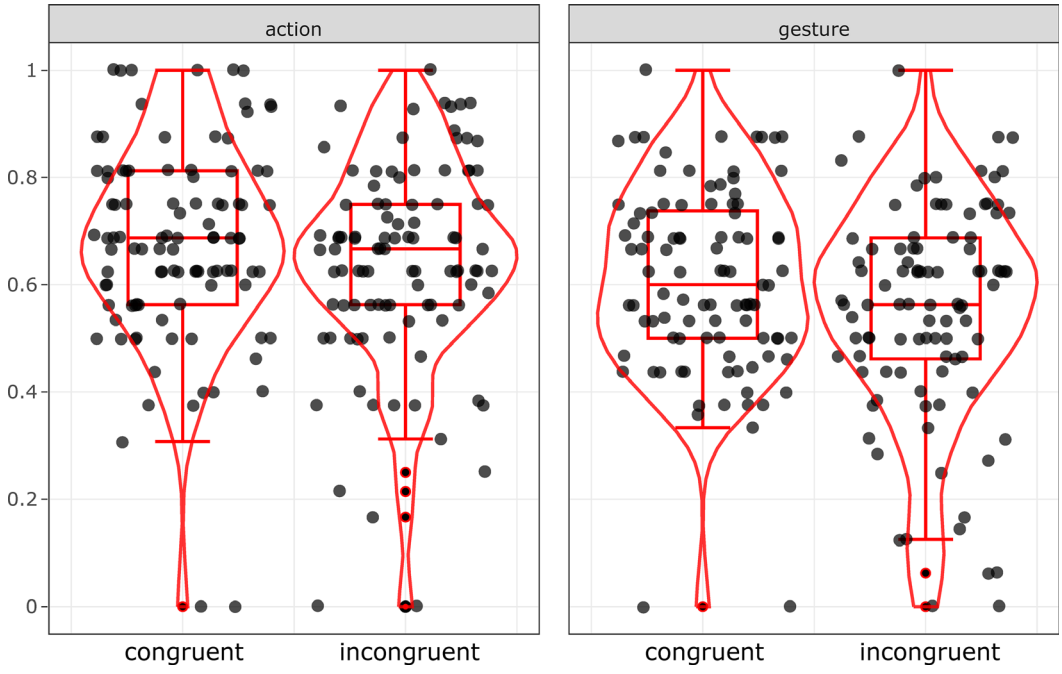

Fig. 3. Jitter-, box-, and violin plot showing the average proportion of trials answered correctly by participants (jitters) and their distribution (box-plot and violin) per condition in Experiment 1. Chance performance in all conditions is 0.50. Perfect memory for where the object actually went is 1.0 in both congruent and incongruent conditions, while perfect use of the handgrasp/end-location rule is 1.0 in the congruent condition and 0 in the incongruent condition.

where each object actually went (regardless of how it was held) would be reflected in a score of 1.0 on both congruent and incongruent trials, while total reliance on the handgrasp/end-location rule to categorize each object would result in a score of 1.0 on congruent trials and a score of 0 on incongruent trials. We examined whether performance was reliably different from chance by running a mixed logistic regression (R package `lme4`; Bates, Maechler, Bolker, & Walker, 2015), with a simple intercept model including random intercepts for participant and object: value ~1 + (1|participant) + (1|object). The intercept was significant, $b = 0.59$, $SE = 0.11$, $z = 5.50$, $p < .001$, demonstrating that the log odds of success are significantly different from zero (log odds of zero correspond to 0.5 event probability which is chance performance). Overall, people were better than chance at indicating whether each object had been placed on the top or bottom. We also ran the equivalent model described above with the data from each condition separately to examine whether performance exceeded chance in each condition. Participants performed reliably above chance in the congruent action ($M = 0.68$, $SE = 0.01$), the incongruent

action ($M = 0.65$, $SE = 0.01$), and the congruent gesture ($M = 0.62$, $SE = 0.01$) conditions, all bs > 0.55, $z > 4.73$, ps < .001. Performance in the incongruent gesture condition ($M = 0.56$, $SE = 0.01$) was not reliably different from chance under our conservative alpha, $b = 0.29$, $z = 2.49$, $p = 0.013$.

We next compared accuracy to indicate the ending location of each object between conditions with a mixed logistic model that included condition (action vs. gesture; between-subject), congruence of the grasp shown in the image (congruent vs. incongruent with the training videos; within-subjects), and the condition × congruence interaction as fixed factors. Participant and object were included as random factors. We initially included three random slopes for object (condition, congruence, and their interaction), and a random slope for participant (congruence), which is the maximal model possible for this mixed design. However, this maximal model did not converge, suggesting that the random effects structure did not adequately fit our data (see Barr, Levy, Scheepers, & Tily, 2013). We dropped the congruence × condition interaction term from the random effects structure associated with object. The final fitted model was accuracy ~ Condition × Congruence + (Congruence|Participant) + (Condition + Congruence|Object). Further, we used centered contrast coding to model the main effects (rather than simple main effects) associated with both Condition and Congruence. Analysis code is available at https://osf.io/de4rm/.

The full results of the model are displayed in Table 2. As predicted, there was a main effect of congruence, such that participants were able to identify the ending location of the object correctly more often when the object was shown being held with the same handgrasp that they had seen demonstrated for the object in the training videos than when the object was shown with the alternative handgrasp, $b = 0.11$, $SE = 0.04$, $z = 2.78$, $p = .005$. This suggests that participants did gain some implicit understanding of the

Table 2
Results of the mixed logistic regression models for performance on the memory task in both experiments

				Random Effects			
	Fixed Effects			By Participants		By Items	
	Parameters	Estimate	z	Variance	Corr	Variance	Corr
Experiment 1	Intercept	0.60	5.48***	0.23	—	0.23	—
	Condition	0.18	3.39***	—	—	0.05	−0.87
	Congruence	0.11	2.78**	0.42	−0.49	0.001	−0.98
	Condition × Congruence	0.02	0.51	—	—	—	—
Experiment 2	Intercept	0.62	6.74***	0.27	—	0.13	—
	Condition	0.19	3.01**	—	—	0.09	−0.80
	Congruence	0.12	2.37*	0.43	−0.66	0.01	−0.21
	Condition × Congruence	0.07	1.34	—	—	—	—

Note. Contrast coding was used to model main effects, and positive estimates represent better performance in the action condition than in the gesture condition, and in the congruent than incongruent condition.
***$p < .001$; **$p < .01$; *$p < .05$.

correspondence between handgrasp and ending location in both the action and the gesture condition. However, contrary to our prediction, there was no condition × congruence interaction, $b = 0.02$, $SE = 0.04$, $z = 0.51$, $p = .61$, suggesting that both action and gesture promote implicit understanding of how to move the objects to the same degree. Although not what we predicted, this finding is noteworthy in light of the final result from the model: a significant effect of condition, such that participants were more likely to accurately remember where the objects went when they had seen the woman actually move the objects than when they had seen the woman gesture about moving the objects, $b = 0.18$, $SE = 0.05$, $z = 3.39$, $p < .001$. Thus, even though gestures were worse than action at promoting memory for where each specific object went, gestures were as good as action at promoting implicit knowledge of the contingency between how to grasp an object and where to put it.

2.2.2. Explicit identification of the handgrasp rule

The likelihood of identifying the handgrasp/end-location rule before the hint was low overall, with just 4% of participants who saw action identifying the rule and 9% of participants who saw gesture identifying the rule. This difference was not significant, $b = 0.79$, $SE = 0.63$, $z = 1.247$, $p = .21$. Following the hint that the woman's rule had something to do with how she moved the objects, more participants (37% in the action condition; 44% in the gesture condition) successfully identified the handgrasp/end-location rule compared to before the hint was given. However, the likelihood of getting the rule was still unaffected by whether participants had seen action or gesture during the training videos, $b = 0.30$, $SE = 0.30$, $z = 0.998$, $p = .32$.

3. Summary

We found that participants' memory for where objects were placed was most accurate when they had seen the woman physically move the objects compared to when they had seen the woman gesture about moving the objects. At the same time, seeing gesture led to just as good implicit understanding of how to move the objects as seeing action, even though participants were largely unable to explicitly state this understanding. Our hypothesis that seeing gesture might promote *better* understanding about how to move the objects than seeing action was not supported.

4. Experiment 2

One limitation of Experiment 1 is that neither the actions nor the gestures were accompanied by speech. Given that some have argued that gestures are privileged over actions when they are produced along with speech (Kelly et al., 2015), it is possible that listeners may learn more from speech-accompanying gestures than from speech-accompanying actions. For example, perhaps the movements of the woman in the gesture condition of

Experiment 1 were not interpreted as gestures with a communicative goal, because she did not provide any information in speech to contextualize the movements. In contrast, in the action condition, where she actually moved the objects, her goal and intention were clearer. Thus, perhaps the actions were easier to remember than the gestures because they were easier for the participants to interpret. To examine this possibility, we conducted Experiment 2 (preregistered at https://osf.io/wuh3e/), in which the actions and gestures were accompanied by speech about where to put each object.

4.1. Method

4.1.1. Participants

Participants were recruited through the online participant pool at Utah Valley University. We initially intended to collect data from a similarly sized sample as in Experiment 1, and we were successful in collecting data from 161 participants before the end of the semester. We excluded data from those who reported experiencing issues with the videos (e.g., the videos did not play, they were too slow to load and did not play in their entirety) or with the audio (e.g., they could not hear anything). The final sample for analysis included 115 participants (74 women; 41 men), with an average age of 21.80 years (SD = 5.49). The majority of the sample (80%) self-identified as White or Caucasian, with 14% identifying as Hispanic or Latinx, 2% as multiracial, 1% as Black, 1% as Asian, and 2% choosing not to disclose their race. The majority of participants in the analyzed sample (n = 112) rated their proficiency with English as a 5 (completely fluent) on a 5-point scale, and 94% reported being exposed to English from birth. As in Experiment 1, participants were randomly assigned to either the gesture (n = 60) or action (n = 55) condition. Participants were compensated with credit toward their research requirement in their introductory psychology course.

4.1.2. Stimuli and procedure

The stimuli and procedure were identical to those used in Experiment 1 (see Section 2.1), except that speech accompanied the woman's actions or gestures in the training videos. Specifically, for each object, the woman began by saying "Put the {name of object}" as she moved her hands from center over to either grasp (in the action condition) or mime grasping (in the gesture condition) the object positioned on the platform to her right. She then finished her sentence by saying "here" as she moved her hands and object (in the action condition) or hands alone (in the gesture condition) toward either the top or bottom shelf. Note that this speech was filmed in the videos originally and had been muted during Experiment 1. Participants spent approximately 12 min completing all stages of the procedure (M = 724.17 s, SD = 490.79).

4.1.3. Data analysis

We followed the same procedures for data coding and analysis as detailed in the description of Experiment 1. As outlined in the preregistration of this second experiment

(https://osf.io/wuh3e/), we adopted a traditional alpha of 0.05 for analysis of the memory task, as it is the task that showed effects in Experiment 1. We maintained the more conservative alpha of 0.01 for all other analyses. This report describes only the memory (implicit understanding) and rule identification (explicit understanding) tasks; the results of the other measures can be found in Appendix S1 and do not change the overall conclusions.

4.2. Results and discussion

4.2.1. Performance on the memory task

As in Experiment 1, we excluded trials for which participants took longer than 5 s to respond or for which Qualtrics recorded a response time of 0 (the participant likely pushed a button before the picture had loaded). This resulted in the loss of 7.5% of the data, and the average number of trials included per participant was 29.6 (out of 32). The accuracy of responses on each trial was coded with the same automated procedure as in Experiment 1 (code available at https://osf.io/cn35p/).

The data are shown in Fig. 4. As in Experiment 1, chance performance in all conditions is .50. Perfect memory for where the objects went regardless of handgrasp is 1.0, and complete reliance on the handgrasp/end-location rule would result in a score of 1.0 on the congruent trials and 0 on the incongruent trials.[3] Performance exceeded 0.5 chance overall, b intercept = 0.59, SE = 0.09, z = 6.62, $p < .001$ as well as individually in the action congruent (M = 0.70, SE = 0.02), action incongruent (M = 0.66, SE = 0.02), and gesture congruent (M = 0.64, SD = 0.02) conditions, bs > 0.63, $z > 5.05$, $p < .001$. As in Experiment 1, performance in the gesture incongruent condition (M = 0.56, SE = 0.02) did not significantly differ from chance, b = 0.25, SE = 0.14, z = 1.80, p = .07.

To compare performance between conditions, we began by fitting the same maximal mixed model attempted in Experiment 1. As in Experiment 1, the maximal model did not converge, so we dropped the random slope for objects associated with the congruence × condition interaction. The final fitted model was thus identical to that used in Experiment 1. We again used contrast coding to model main effects, and the analysis code can be found at https://osf.io/vuwbr/.

The full results of the analysis are displayed in Table 2. The findings mirror those from Experiment 1. There was a main effect of condition, such that participants who saw the woman actually move the objects had better memory for their ending location than participants who saw the woman gesture about moving the objects, b = 0.19, SE = 0.06, $z = -3.01$, p = .002. There was also an effect of congruence, b = 0.12, SE = 0.05, $z = -2.37$, p = .018, such that participants had better memory for the location of each object when they saw the woman holding it in the same way she had grasped (or pretended to grasp) it in the training videos than when they saw her holding it with the alternative grasp. There was no condition × congruence interaction, b = 0.07, SE = 0.05, $z = -1.34$, p = .18. Thus, as in Experiment 1, participants' memory for where the objects went was strongest when they had seen action, although gesture did lead to a congruence effect of similar magnitude as that in the action condition. This suggests that gesture and

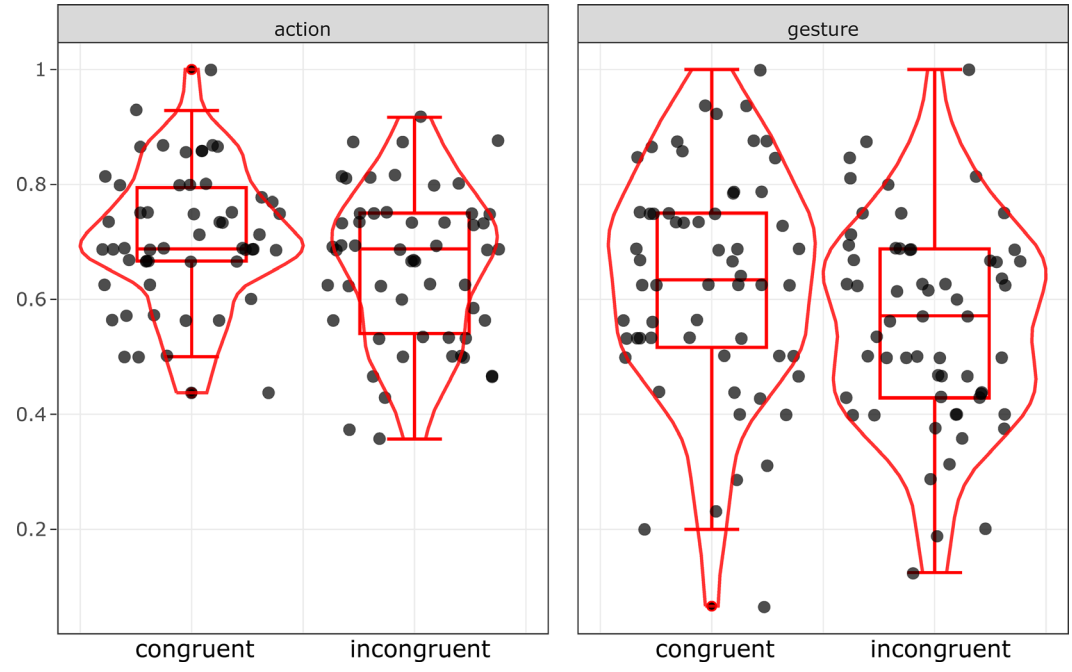

Fig. 4. Jitter-, box-, and violin plot showing the average proportion of trials answered correctly by participants (jitters) and their distribution (box-plot and violin) per condition in Experiment 2. Chance performance in all conditions is 0.50. Perfect memory for where the object actually went is 1.0 in both congruent and incongruent conditions, while perfect use of the handgrasp/end-location rule is 1.0 in the congruent condition and 0 in the incongruent condition.

action promoted implicit understanding of the contingency between a particular handgrasp and a particular end location.

4.2.2. Explicit identification of the handgrasp rule

As in Experiment 1, very few participants identified the handgrasp rule before the hint, and it did not matter whether they had seen action (with 9% stating the rule) or gesture (12% stating the rule), $b = 0.28$, $SE = 0.62$, $z = 0.45$, $p = .65$. After the hint, 50% of participants identified the rule, and this was unaffected by whether participants had seen action (55%) or gesture (45%) in the training videos, $b = -0.38$, $SE = 0.375$, $z = -1.02$, $p = .31$.

4.2.3. Summary

The results mirror those from Experiment 1 and suggest that seeing actions leads to better memory than seeing gestures for the outcomes of the actions (i.e., where the

objects end up), regardless of whether there is accompanying speech. However, at the same time, seeing gestures is as effective as seeing actions at promoting implicit understanding of the sensorimotor regularities involved in moving the objects, even though participants were largely unable to explicitly state the rule about those regularities.

5. General discussion

Across two experiments, we found a consistent pattern of results showing that (a) actions are better than gestures at promoting memory for the end state of an action, but (b) both gestures and actions can promote an implicit understanding of movement patterns to the same degree. Although neither of these findings is exactly what we predicted, taken together, they do suggest that gestures are better at highlighting the particulars of a movement pattern such as how an object should be moved (where they do not differ from action) than they are at highlighting end state information such as where the object should be placed (where they are worse than action).

We found that observers have better memory for where objects are placed when they have seen a woman actually move the objects to their end location than when they have seen her gesture about where to move the objects. Although we did not explicitly predict this finding, it does align with the proposal that actions are interpreted as goal-directed and direct observers' attention to objects and end states (e.g., Novack et al., 2016; Schachner & Carey, 2013; Woodward, 1998). Further, in the paradigm used here, it must be noted that the action condition contained an image of each object in its ending location that was not present in the gesture condition. While the action training videos ended with the object positioned on either the top or bottom shelf, the objects in the gesture training videos never left the starting platform. Thus, the most obvious explanation for the superior memory of participants in the action condition is that people have better memory for things they actually saw (e.g., the actual object in its end location) than for things that they only imagined (e.g., imagining the object moving from the starting platform to the indicated end location). The finding also dovetails with some previous reports that actions are processed more easily than gestures (e.g., Kelly et al., 2015), perhaps because gestures require additional cognitive effort to imagine their referent.

If we assume that the superior memory for the end state of the object in the action condition is due to participants seeing (rather than only imagining) the object in its end state, a similar perception-based explanation can be applied to the congruency effect in the action condition. Because participants had seen the object being held with a particular hand grasp during training, cueing that particular hand grasp during recall improved performance, as this was the context under which the object's location was learned. Essentially, making the perceptual context at retrieval as similar as possible to what it was at encoding should (and did) help memory. But this is precisely why the congruency effect in the gesture condition becomes especially interesting. In the gesture condition, the object was never seen being held with either type of hand grasp during training. The still images shown in the memory phase of the gesture condition were all novel images,

regardless of whether they were congruent or incongruent, as the participants in the gesture condition had never seen the woman actually holding the objects in any manner during the training phase. Thus, the congruency effect in the gesture condition cannot be explained by people responding more accurately to images they have seen before than to images they have not seen before.

Instead, the congruency effect in the gesture condition suggests that people had gained implicit knowledge of the contingency between handgrasp and ending location from watching the woman gesture. Even though seeing gesture did not help participants visualize where the object ended up as well as seeing an action that placed the object there, gestures were effective at highlighting the movement patterns involved in moving the objects. It appears then that gesture can signal a sensorimotor regularity about how to interact with objects, even though the objects are never *actually grasped* nor *actually moved*. This finding reveals a quite remarkable power of representation through gesture, and it lends empirical support to the idea that gestures provide effective demonstrations of how to complete an action on an object, even when the object is not physically being manipulated (Cataldo et al., 2018; Gärdenfors, 2017). As such, gestures are a cheap and portable resource for teaching sensorimotor associations that do not require the presence of objects in the same way actions do.

We predicted that knowledge of the contingency between handgrasp and ending location might be promoted more strongly by gesture than by action, but this hypothesis was not supported. This hypothesis was motivated largely by the developmental literature, which suggests that children have better memory for how a movement is performed when this movement is represented through gesture than when it is shown through action (Aussems & Kita, 2017) and that children are better able to generalize the manner of a verb that has been taught through gesture than through action (Mumford & Kita, 2014). It is possible that the relative influence of gesture and action changes in adulthood, so that gestures are no longer superior to action in this way for adults. Alternatively, it is possible that gesture is actually superior to action at showing the particulars of movement patterns, but we did not detect the difference in our task. For example, it would be interesting to use a more reliable measure of reaction time in future work, as it is possible that gesture may affect the relative reaction time to make a decision about where the object should go in the congruent versus incongruent conditions, even without affecting relative accuracy.

Finally, although we found evidence in the memory task for implicit understanding of the handgrasp/end-location rule, participants seemed largely unaware of the relationship. Only 10% of participants across the two experiments were able to explicitly state the rule before any hint was given, and only about half could do it even after receiving a hint that the rule had to do with how she moved the objects. One possibility is that the participants in our study were not particularly motivated doing this study online; it would be worthwhile to see whether participants would be more likely to get the rule in a laboratory setting. Nonetheless, we suspect that explicit recognition of the rule was difficult because humans have a natural tendency to focus on perceptual properties (e.g., an object's color or shape), taxonomic similarities (i.e., kind of object), or thematic similarities (i.e., how

objects are relationally similar to each other, because of related functions or contexts in which they would appear) when forming categories (e.g., Golinkoff, Shuff-Bailey, Olguin, & Ruan, 1995; Imai, Gentner, & Uchida, 1994; Landau, Smith, & Jones, 1988; Murphy, 2001). Even after shifting away from a bias to focus on surface-level properties like shape (Landau et al., 1988), older children and adults typically focus on how objects can be grouped based on their type or function (Berger & Donnadieu, 2006; Imai et al., 1994; Murphy, 2001). Given that the 16 items in our study were familiar objects to participants, it is perhaps not surprising that participants' first attempt at grouping them was based on what they knew about the properties and functions of the objects. In contrast, the way in which they were moved introduced a novel rule of categorization that does not align with our natural ways of sorting objects. Perhaps if objects had all been novel, participants would have been more likely to see handgrasp/end location as an explicitly important rule by which to categorize the objects.

Regardless of why so many participants were unable to explicitly state the rule, the fact that they could not suggests that gestures and actions primarily affected implicit understanding of the relationship between handgrasp and end location. This aligns with previous reports from clinical populations (e.g., Hilverman et al., 2018; Klooster et al., 2014) that gestures affect learning through implicit processes. Of course, once implicit knowledge has been gained, that knowledge can become explicit. It would be interesting to use this paradigm in future studies to examine whether some types of experiences, such as producing the actions or gestures oneself (e.g., Hilverman et al., 2018) or sleeping (e.g., Cook et al., 2013; Wilhelm et al., 2013), might help participants explicitly understand the rule. It could also be interesting to explore whether speakers of languages that prioritize manner and path information differently than English (see Talmy, 1983) might understand the rule more readily than the English speakers used here.

In sum, much previous research has compared what people learn from speech with gesture to what people learn when they hear the same speech without gesture (e.g., Rueckert, Church, Avila, & Trejo, 2017). The evidence from such studies is clear that listeners learn more with gesture than they learn without (Dargue et al., 2019; Hostetter, 2011). However, to begin to understand how gestures might have their communicative power, we argue that it is useful to compare them to other sorts of nonverbal information that speakers might use, such as pictures or actions on objects. In this study, we have shown that gestures and actions on objects are not equivalent—if the goal is to emphasize end state information, action is likely best. But if the goal is to show movement patterns, gestures can work just as well.

Open Research badge

This article has earned Open Data badge. Data are available at https://osf.io/v35u6/.

Notes

1. Of course, it is possible that there is some other rule we have not thought of that could differentiate the objects in Set A from those in Set B. However, we checked every rule that our participants generated in the experiment, and none of them works to categorize all instances correctly. If there is some other possible rule to distinguish between the two sets of objects, our participants were not aware of it.
2. Although we encouraged participants to respond as quickly as possible, the reaction time data recorded in the Qualtrics platform are not precise enough for meaningful analyses.
3. One participant in the gesture condition of Experiment 2 showed a complete reversal of the rule they had seen when responding to the memory trials. That is, they always responded that two-handed objects should go up and one-handed objects should go down, when the training videos they had seen depicted the reverse. As a result, they scored 0 in the congruent condition and 1 in the incongruent condition. It is unclear why they responded in this way, as they did not mention handgrasp at all in any of their attempts to state the rule explicitly. Because of their unusual pattern, we did run the analysis without their data included, and it does not change the significance of the patterns observed. The results reported here include this person's data.

References

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

Aussems, S., & Kita, S. (2017). Seeing iconic gestures while encoding events facilitates children's memory of these events. *Child Development*, *90*, 1123–1137. https://doi.org/10.1111/cdev.12988

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*, 255–278. https://doi.org/10.1016/j.jml.2012.11.001

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*, 1–48. https://doi.org/10.18637/jss.v067.i01

Berger, C., & Donnadieu, S. (2006). Categorization by schema relations and perceptual similarity in 5-year-olds and adults: A study of vision and audition. *Journal of Experimental Child Psychology*, *93*, 2304–2321. https://doi.org/10.1016/j.jecp.2005.10.001

Borghi, A. M., & Riggio, L. (2015). Stable and variable affordances are both automatic and flexible. *Frontiers in Human Neuroscience*, *9*, 351.

Buresh, J. S., & Woodward, A. (2007). Infants track action goals within and across agents. *Cognition*, *104*, 287–314. https://doi.org/10.1016/j.cognition.2006.07.001

Cataldo, D. M., Migliano, A. B., & Vinicius, L. (2018). Speech, stone tool-making and the evolution of language. *PLoS ONE*, *13*(1), e0191071. https://doi.org/10.1371/journal.pone.0191071

Cook, S. W., Duffy, R. G., & Fenn, K. M. (2013). Consolidation and transfer of learning after observing hand gesture. *Child Development*, *84*(6), 1863–1871. https://doi.org/10.1111/cdev.12097

Dargue, N., Sweller, N., & Jones, M. P. (2019). When our hands help us understand: A meta-analysis into the effects of gesture on comprehension. *Psychological Bulletin*, *145*, 765–784. https://doi.org/10.1037/bul0000202

Dienes, Z., & Berry, D. (1997). Implicit learning: Below the subjective threshold. *Psychonomic Bulletin & Review*, *4*(1), 3–23. https://doi.org/10.3758/BF03210769

Gärdenfors, P. (2017). Demonstration and pantomime in the evolution of teaching. *Frontiers in Psychology*, *8*, 415. https://doi.org/10.3389/fpsyg.2017.00415.

Golinkoff, R. M., Shuff-Bailey, M., Olguin, R., & Ruan, W. (1995). Young children extend novel words at the basic level: Evidence for the principle of categorical scope. *Developmental Psychology*, *31*, 494–507. https://doi.org/10.1037/0012-1649.31.3.494

Hilverman, C., Cook, S. W., & Duff, M. C. (2018). Hand gestures support word learning in patients with hippocampal amnesia. *Hippocampus*, *28*, 406–415. https://doi.org/10.1002/hipo.22840

Hostetter, A. B. (2011). When do gestures communicate? *A meta-analysis. Psychological Bulletin*, *137*(2), 297–315. https://doi.org/10.1037/a0022128

Imai, M., Gentner, D., & Uchida, N. (1994). Children's theories of word meaning: The role of shape similarity in early acquisition. *Cognitive Development*, *1*, 45–75. https://doi.org/10.1016/0885-2014(94)90019-1

Kelly, S., & Church, R. B. (1997). Can children detect conceptual information conveyed through other children's nonverbal behaviors? *Cognition and Instruction*, *15*, 107–134. https://doi.org/10.1207/s1532690xci1501_4

Kelly, S., Healey, M., Özyürek, A., & Holler, J. (2015). The processing of speech, gesture, and action during language comprehension. *Psychonomic Bulletin & Review*, *22*(2), 517–523. https://doi.org/10.3758/s13423-014-0681-7

Kita, S., Alibali, M. W., & Chu, M. (2017). How do gestures influence thinking and speaking? The gesture-for-conceptualization hypothesis. *Psychological Review*, *124*, 245–266. https://doi.org/10.1037/rev0000059

Klooster, N. B., Cook, S. W., Uc, E. Y., & Duff, M. C. (2014). Gestures make memories, but what kind? Patients with impaired procedural memory display disruptions in gesture production and comprehension. *Frontiers in Human Neuroscience*, *8*, 1054. https://doi.org/10.3389/fnhum.2014.01054

Landau, B., Smith, L. B., & Jones, S. S. (1988). The importance of shape in early lexical learning. *Cognitive Development*, *3*, 299–321. https://doi.org/10.1016/0885-2014(88)90014-7

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

Mumford, K. H., & Kita, S. (2014). Children use gesture to interpret novel verb meanings. *Child Development*, *85*(3), 1181–1189. https://doi.org/10.1111/cdev.12188

Murphy, G. L. (2001). Causes of taxonomic sorting by adults: A test of the thematic-to-taxonomic shift. *Psychonomic Bulletin & Review*, *8*, 834–839. https://doi.org/10.3758/bf03196225

Novack, M. A., Wakefield, E. M., & Goldin-Meadow, S. (2016). What makes a movement a gesture? *Cognition*, *146*, 339–348. https://doi.org/10.1016/j.cognition.2015.10.014

Reber, A. S. (1989). Implicit learning and tacit knowledge. *Journal of Experimental Psychology: General*, *118*(3), 219–235. https://doi.org/10.1037/0096-3445.118.3.219

Rueckert, L., Church, R. B., Avila, A., & Trejo, T. (2017). Gesture enhances learning of a complex statistical concept. *Cognitive Research: Principles and Implications*. Article 2. https://doi.org/10.1186/s41235-016-0036-1

Schachner, A., & Carey, S. (2013). Reasoning about "irrational" actions: When intentional movements cannot be explained, the movements themselves are seen as the goal. *Cognition*, *129*, 309–327. https://doi.org/10.1016/j.cognition.2013.07.006

Singer, M. A., & Goldin-Meadow, S. (2005). Children learn when their teacher's gestures and speech differ. *Psychological Science*, *16*(2), 85–89.

Talmy, L. (1983). How language structures space. In H. Pick & L. Acredelo (Eds.), *Spatial orientations: Theory, research, and application* (pp. 225–282). New York: Plenum Press.

van Wermeskerken, M., Fijan, N., Eielts, C., & Pouw, W. T. J. L. (2016). Observation of depictive versus tracing gestures selectively aids verbal versus visual–spatial learning in primary school children. *Applied Cognitive Psychology*, *30*(5), 806–814. https://doi.org/10.1002/acp.3256

Wakefield, E., Hall, C., James, K. H., & Goldin-Meadow, S. (2018). Gesture for generalization: Gesture facilitates flexible learning of words for actions on objects. *Developmental Science*, *21*, e12658. https://doi.org/10.1111/desc.12656

Wakefield, E., Novack, M. A., Congdon, E. L., Franconeri, S., & Goldin-Meadow, S. (2018). Gesture helps learners learn, but not merely by guiding visual attention. *Developmental Science*, *21*, e12664. https://doi.org/10.1111/desc.12664

Wilhelm, I., Rose, M., Imhof, K. I., Rasch, B., Buchel, C., & Born, J. (2013). The sleeping child outplays the adult's capacity to convert implicit into explicit knowledge. *Nature Neuroscience*, *16*(4), 391–393. https://doi.org/10.1038/nn.3343

Woodward, A. (1998). Infants selectively encode the goal object of an actor's reach. *Cognition*, *69*, 1–34. https://doi.org/10.1016/S0010-0277(98)00058-4

Supporting Information

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

Appendix S1. Additional Analyses.