



HHS Public Access

Author manuscript

Cognition. Author manuscript; available in PMC 2017 January 01.

Published in final edited form as:

Cognition. 2016 January ; 146: 339–348. doi:10.1016/j.cognition.2015.10.014.

What makes a movement a gesture? ★

Miriam A. Novack^{#*}, Elizabeth M. Wakefield[#], and Susan Goldin-Meadow

The University of Chicago, United States

[#] These authors contributed equally to this work.

Abstract

Theories of how adults interpret the actions of others have focused on the goals and intentions of actors engaged in object-directed actions. Recent research has challenged this assumption, and shown that movements are often interpreted as being for their own sake (Schachner & Carey, 2013). Here we postulate a third interpretation of movement—movement that *represents* action, but does not literally act on objects in the world. These movements are gestures. In this paper, we describe a framework for predicting when movements are likely to be seen as representations. In Study 1, adults described one of three scenes: (1) an actor moving objects, (2) an actor moving her hands in the presence of objects (but not touching them) or (3) an actor moving her hands in the absence of objects. Participants systematically described the movements as depicting an object-directed action when the actor moved objects, and favored describing the movements as depicting movement for its own sake when the actor produced the same movements in the absence of objects. However, participants favored describing the movements as *representations* when the actor produced the movements near, but not on, the objects. Study 2 explored two additional features—the form of an actor’s hands and the presence of speech-like sounds—to test the effect of context on observers’ classification of movement as representational. When movements are seen as representations, they have the power to influence communication, learning, and cognition in ways that movement for its own sake does not. By incorporating representational gesture into our framework for movement analysis, we take an important step towards developing a more cohesive understanding of action-interpretation.

Keywords

Gesture; Action-understanding; Representational movement

1. Introduction

When a runner grabs a water bottle (or a pint of ice cream) after a long run, we quickly interpret his action as being goal-directed and intentional. We do not puzzle over the specific

[★]This work was supported by grants from NICHD (R01-HD47450) and NSF (BCS-0925595) to Goldin-Meadow, a grant from the Institute of Education Sciences (R305 B090025) to S. Raudenbush in support of Novack, a Training Grant from NICHD (T32-HD07475) to L.B. Smith in support of Wakefield, and by the Spatial Intelligence and Learning Center funded by NSF (SBE 0541957, Goldin-Meadow is a PI).

^{*} Corresponding author at: Department of Psychology, The University of Chicago, Chicago, IL 60637, United States. mnovack1@uchicago.edu (M.A. Novack).

arm movements used to acquire the water or ice cream; we understand that his actions were performed to reach a goal: to quench his thirst or reward himself with a treat. This example illustrates a skill found in both adults and children—the ability to interpret the actions and intentions of others as goal-directed: actions that are guided by top-down, hierarchical cognitive processes (e.g., Baldwin & Baird, 2001; Bower & Rinck, 1999; Searle, 1980; Trabasso & Nickels, 1992; Zacks, Tversky, & Iyer, 2001).

The literature on action understanding typically assumes that, when an actor performs a movement, the actor's goal is not the movement itself, but rather the impact that the movement has on the surrounding world. Recently, Schachner and Carey (2013) challenged this assumption as overly simplistic by demonstrating that, under certain conditions, movements can be interpreted, not as a means to an external goal, but instead as a goal unto itself—movement for the sake of movement. Here we broaden the investigation of action understanding to include a third type of movement. We examine the conditions under which movement² is interpreted not as movement directed toward an object, nor as movement performed for its own sake, but rather as movement that *represents* other types of actions. This third type of movement is *gesture*.

A foundational body of research demonstrates that humans interpret actions in terms of the actor's intentions and goals (e.g., Baldwin & Baird, 2001; Bower & Rinck, 1999; Searle, 1980; Trabasso & Nickels, 1992; Zacks et al., 2001). For example, adults naturally parse continuous action into smaller, goal-directed segments (Newtson, 1976), and when asked to describe ambiguous scenes, adults and 5-year-old children will create a goal-based explanation for an actor's actions (Trabasso, Stein, Rodkin, Munger, & Baughn, 1992). The ability to interpret object-directed actions in terms of an actor's goals arises in the first year of life—infants as young as 6-months notice when an actor uses the same movements to reach for a new goal, but not when the actor uses new movements to reach the same goal (Buresh & Woodward, 2007; Woodward, 1998). In other words, infants consider a change in what is being reached for as novel and worth attending to, but do not consider a change in the movements made to achieve the same goal as noteworthy or surprising. At 10–11 months, children are surprised when an actor's reaching movement stops abruptly before an object is grasped, but not when the actor's reaching movement stops after the object has been grasped (Baldwin, Baird, Saylor, & Clark, 2001). There is thus ample evidence that both adults and children interpret movements as goal-directed. An important caveat, however, is that much of the relevant research has been done on movements directed toward objects, with an emphasis on external goals.

Recently, Schachner and Carey (2013) broadened the scope of these studies to include movement performed in the absence of objects, and showed that, under these circumstances, adults tend to interpret the movement as having its own intrinsic goal, as movement for its own sake. In the first in a set of experiments, participants were asked to interpret videos in which a character either moved objects from one location to another (*Objects Present* condition) or made the same movements without the objects present (*Objects Absent*

²Here, we use the word movement to refer only to intentional actions. Accidental or unintentional actions (e.g., accidentally knocking a glass off a table; moving up and down on a merry-go-round) are beyond the scope of this paper.

condition). Participants described the movements in terms of external goals when the objects were present (e.g., his intention was to sort the colored balls), but in terms of movement-based goals when the objects were absent (e.g., his intention was to jump into the air and move to the left and right). In another experiment, Schachner and Carey (2013) showed that participants also attribute movement-based goals to an actor producing inefficient movements. The actor's movements were either inefficient (moving toward the target and then away from it) or efficient (moving toward the target), and participants were more likely to provide movement-based descriptions for inefficient than for efficient movements. The traditionally held view that humans interpret actions in terms of external goals thus seems to be too narrow. In certain circumstances, adults will see the intent of an action as the completion of the action itself—movement for the sake of movement.

Into this two-dimensional perspective of how humans interpret action, we propose another dimension of intentional, goal-directed movement—gesture. Gestures are movements of the hands that accompany speech and communicate information to listeners (Kendon, 1994). Although researchers have created elaborate coding systems for identifying, describing, and interpreting the meaning of gestural movements (e.g., Church, Kelly, & Wakefield, 2015; Kendon, 2004; McNeill, 1992), and even naïve observers, who are not trained gesture coders, are able to reliably interpret gestural movements (Goldin-Meadow & Sandhofer, 1999), there has not been a systematic investigation of the circumstances under which observers interpret movement as a representation (i.e., as gesture), as opposed to goal-directed or for its own sake. Gesture is unlike object-directed movement, whose goal is to achieve some change in the world, and unlike movement for its own sake, whose goal is to produce the movement itself. In contrast to both of these goals, the purpose of gesture is to reference or represent other movements, objects, or even abstract ideas. Importantly this, can include movements that represent *either* external or movement-based goals. In other words, gesture does not depict a change in the world (e.g., opening a jar by twisting it), or display patterns of movement (e.g., performing steps in a dance), but instead *represents* movement that could change the world (e.g., a gesture showing how the jar could be twisted open) or *represents* movement that stands on its own (e.g., a gesture showing how the dance should be performed). Importantly, observers respond differently to gesture than to other types of movements (Dick, Goldin-Meadow, Hasson, Skipper, & Small, 2009; Kelly, Healy, Ozyurek, & Holler, 2014), and these differences in response can have an impact on thinking and learning (Novack, Congdon, Hemani-Lopez, & Goldin-Meadow, 2014; Trofater, Kontra, Beilock, & Goldin-Meadow, 2014). Our goal is to determine the conditions under which a distinction between gesture and other types of movement is made, and to explore how those conditions contribute to a top-down categorization of movement as gesture.

We begin by noting that a movement has the capacity to be seen as a gesture if it does *not* cause effects on the external environment. As a result, any empty handed-movement is a candidate for being interpreted as a gesture. Our first prediction, then, is that hand movements that interact with objects will likely be interpreted as having external goals, whereas hand movements that occur *off* objects have the potential to be seen either as movements produced for their own sake, or as movements that represent.³

The next question is how to determine whether an empty-handed movement is a movement meant for its own sake, or a movement meant to represent. We suggest that humans have a bias to interpret movement as more than movement for its own sake. In fact, Schachner and Carey (2013) found that almost onethird of participants who saw a character move in empty space *still* attempted to describe his movement in terms of external-goals. Seeing empty-handed movement as movement for its own sake may be a default that is activated only when observers are unable to see the movement as anything else. If there are contextual cues that could confer meaning to an empty-handed movement, observers are likely to exploit those cues and interpret the movement as meaningful and, as a consequence, seek something that the movement could represent. Our second prediction, then, is that empty-handed movements shift from being seen as movement for their own sake to movements that represent when there is information in the context that allows observers to interpret the movements as meaningful.

Taken together, these predictions form a framework in which hand movements are seen as having external goals, movement-based goals, or representational goals depending on (1) whether the hands interact with objects and, if not, (2) whether there is information in the context that makes the movements appear meaningful to the observer. Finally, it is important to note that this framework focuses on the *hands*. Hands may be special in their ability to both carry out goals *and* communicate or represent, and may therefore be a special cue for seeing movement as representational.

We explore this framework across two studies. In the first study, we test the prediction that movements that do not interact with objects (i.e., empty-handed movements) are candidates for having representational goals if there is information in the context that makes them seem meaningful to the observer. We extend Schachner and Carey's (2013) paradigm to include a condition that meets these requirements and thus is likely to be interpreted as gesture. Schachner and Carey gave observers two types of scenes to interpret, one in which an actor acts directly on an object (to which observers attributed external-goals), and another in which the actor performs the same movements but without the object present (to which observers attributed movement-based goals). We add a third type of scene—one in which the actor performs the same movements but over (not directly on) the objects, which are present. According to our framework, we expect observers to have a bias to see empty-handed movements as meaningful, particularly if there is information in the context (in this case, the presence of objects) that supports attributing meaning to the movement. Observers should therefore interpret movement over, but not directly on, an object as a *representation* of a goal-directed action on the object, that is, as a gesture. In the second study, we extend our second prediction, expanding the context to include additional cues to meaning. We ask whether these additional cues influence the observer's inclination to interpret empty-handed movement in terms of movement-based goals or representational goals. We predict that contextual cues to meaning—the presence of objects, the presence of speech, the form of the

³In some cases, an empty-handed movement can be interpreted as having an external goal; for example, when an observer thinks an agent is trying to complete an action on an object but the agent fails to make contact with the object (i.e., the agent performs an incomplete action), perhaps because the object is out of reach or is behind a barrier.

movement itself—will increase an observer’s inclination to see an empty-handed movement as a representational gesture.

2. Study 1

As just described, Schachner and Carey (2013) held the movements performed by an actor constant, but varied the presence or absence of objects. When objects were present and acted on, observers systematically believed that the goal of the actor was to move the objects (i.e., an external goal). When objects were absent and the actor produced the same movements in the air rather than on an object, half of the observers attributed a movement-based goal to the actor (i.e., the movement itself was the intended outcome). In Study 1, we add a third condition—the objects are present but are not acted upon—and hypothesize that movements produced in this condition are likely to be interpreted as a gesture representing action.

Study 1 thus contained three conditions, the first two modeled after Schachner and Carey (2013): (1) *Action on Objects*, in which an actor directly manipulates objects; (2) *Action with Objects Absent*, in which an actor performs the same movements, but without the objects present; (3) *Action off Objects with Objects Present*, in which an actor performs the same movements over (not directly on) the objects, which are present. We expect to replicate Schachner and Carey’s findings in the first two conditions: Observers will describe movements in the *Action on Objects* condition in terms of external goals (i.e., “she organized objects”), but will be more likely to describe movements in the *Action with Objects Absent* condition in terms of movement-based goals (i.e., “she moved her hands to the right and left”).

The more interesting question to us is how participants will interpret movements in the third condition. In the *Action off Objects with Objects Present* condition, an actor performs movements in the presence of objects, but does not use her movements to manipulate those objects. Following Schachner and Carey (2013), we might expect that a movement produced in the presence of objects would be interpreted as having a movement-based goal if the movements are seen as intentional yet inefficient. However, if the presence of objects provides context that invites richer interpretation, the movements might instead be seen as representing an external goal.

2.1. Method

2.1.1. Participants—120 adult English-speaking residents of the United States (63 females, 55 males, 2 unreported; 40 participants in each of the 3 conditions) participated in the experiment via Amazon Mechanical Turk (<https://www.mturk.com>), a website through which individuals can complete tasks for small amounts of compensation (see Crump, McDonnell, & Gureckis, 2013 for validation of AMT for experimental studies). All participants were required to have had at least 95% of their previous work on Amazon Mechanical Turk judged as acceptable, and were required to have previously completed at least 100 Mturk studies. An additional 18 participants were excluded if they provided inappropriate answers, suggesting that they had not watched the video stimulus ($n = 2$), or if they completed the experiment multiple times ($n = 16$). Participants were all over 18 years of age, and were ethnically diverse: White ($n = 92$); Black ($n = 11$), Asian ($n = 8$); Native

American ($n = 2$); more than one race ($n = 3$); unreported ($n = 4$). The task took just under 4 min to complete on average, and participants were compensated \$0.25 for their time.

2.1.2. Stimuli—Video stimuli showed the torso of a woman standing in front of a table. Her chest, arms, and hands were visible, but her face was not. All videos were 10 s long and displayed the actor producing movements with her hands. In two conditions (Action on Objects; Action off Objects with Objects Present), four balls (two orange and two blue), as well as two boxes (one orange and one blue), sat on the table in front of the woman. In the third condition (*Action with Objects Absent*), no objects were present on the table. Each video is described in detail below (see Fig. 1 for still frames taken from the videos).

2.1.2.1. Action on Objects: A woman picks up each of the balls on the table in front of her and places them in the color-matched boxes (see Fig. 1a). Specifically, she picks up the inner blue ball with her left hand, and places it in the blue box on her left; then she picks up the inner orange ball with her right hand and places it in the orange box on her right. These actions are repeated with the outer two balls.

2.1.2.2. Action off Objects with Objects Present: A woman produces the same movements in the Action on Objects condition but over the objects, not on them (see Fig. 1b). More specifically, the woman maintains the hand shape necessary to grasp the balls (i.e., a palm down C-shape) and moves her hands following the trajectory she took when actually moving the objects (i.e., from one ball to the box, from another ball to the other box). As the woman never actually touches the balls or boxes, there is no change in the location of the balls during the video.

2.1.2.3. Action with Objects Absent: A woman produces the same movements as in the other two conditions, but without any of the objects present (see Fig. 1c). The woman maintains the same hand shape and trajectory as in the first two conditions.

To control for timing across videos, the *Action on Objects* condition was recorded first. The woman synchronized her movements to a metronome set to 80 beats per minute. After this video was filmed, she recorded a co-speech narrative (“I put one here, another in this box, the orange one in this one, and the last one in this place”) that she synchronized to her movements in the *Action on Objects* video. The narrative was then played back to her as she performed the movements in the other two conditions to ensure that timing would match across videos. Audio was not included in any of the videos.

2.1.3. Procedure—The study was conducted through Amazon Mechanical Turk. Participants were randomly assigned to one of the three conditions. They were told that they would watch a 10-s movie and be asked questions about what they saw. Each participant saw one of the three videos and answered the free-response question, “What happened in the scene?” Finally, participants answered optional demographic questions.

2.1.4. Coding and reliability—Free-responses were classified into one of 3 goal-based categories—External Goal, Movement-based Goal, Representational Goal. Although the prompt “What happened in the scene?” does not specifically probe the intentions of the

agent, the pragmatics of the question, combined with the fact that the agent was producing movements of her own volition, should invite descriptions of goals and intentions (see Schachner & Carey, 2013 for discussion of assumptions about intentionality). A small number of responses ($n = 1$, 0.01% of the total responses) were considered Uncodable. The codes are described below:

1. *External Goal*: The movie is described in terms of actions completed on objects, with a focus on movement of objects, rather than movement of hands (e.g., “a person put blue balls in blue box and orange balls in an orange box”; “balls were placed in boxes”; “the person picked up colored balls one at a time and placed them in color coordinated boxes.”).
2. *Movement-based Goal*: The movie is described in terms of low-level spatiotemporal movements without mentioning a higher-level goal – the description is focused on the movement of the hands themselves (e.g., “a woman waved her hand over two blue spheres and a blue box”, “waving her hand, alternating sides”, “a person was moving hands”).
3. *Representational Goal*: The movie is described in terms of either (a) movements representing external goals (e.g., “she demonstrated how to put balls in boxes”; “she pretended to sort objects”; “gestured about moving colored spheres”), or (b) movements representing movement-based goals (e.g., “the actress was telling you how to perform the steps of a dance and where to put your feet”).
4. *Uncodable*: The movie is described without mentioning movement at all (e.g., “there was a woman and some boxes and balls”).

Two researchers assigned a single code to all responses, and were blind to the condition of each participant. Coders agreed on 93% of trials, $K = 0.90$. Any disagreements were discussed between the coders and resolved.

2.2. Results

Fig. 2 presents the proportion of participants whose responses fell into each of the three primary coding categories. Because the three response codes were not independent, we ran log-linear poisson models, partitioning the 3 (Condition) \times 3 (Response Code) contingency table to conduct three independent analyses, with the aim of determining: (1) whether a different pattern of responses was given across conditions, and (2) whether empty-handed conditions elicited a different number of Movement vs. Representational goal responses.

To begin, we collapsed across the *Action off Objects with Objects Present* and *Actions with Objects Absent* conditions, and asked whether there was a different pattern of responses between these combined, empty-handed conditions, and the *Action on Objects* condition, creating a 2 (Condition) \times 3 (Response) contingency table. Our model revealed that responses in the *Action on Objects* condition were significantly different from responses in the conditions in which empty-handed movements were observed (χ^2 , 2 = 120.98, $p < .001$).

Next, we considered whether the pattern of responses was different across the two, empty-handed conditions. We directly compared responses elicited in the *Action off Objects with*

Objects Present and *Actions with Objects Absent* conditions, creating a 2 (Condition) \times 3 (Response) contingency table, which did not include data from the *Action on Objects* condition. Again, our model revealed a significant difference between conditions ($\chi^2, 2 = 14.73, p < .001$).

Note that external goal responses were extremely rare in both empty-handed movement conditions (3 external goal responses in the *Action off Objects with Objects Present* and 0 in the *Actions with Objects Absent* condition). The difference between the two empty-handed conditions must then have been driven by different rates of movement-based versus representational goal responses. Indeed, if we consider the proportion of each of these response types, we find that, in the *Action off Objects with Objects Present* condition, representational goals were the predominant response (68%), whereas movement-based goals were the dominant response (63%) in the *Actions with Objects Absent* condition. A 2 (Condition) \times 2 (Response) chi-square, comparing the number of movement-based and representational goal responses elicited in the *Action off Objects with Objects Present* versus *Actions with Objects Absent* conditions, confirmed this difference as significant ($\chi^2, 1 = 10.54, p < .001$).

Our results provide support for a framework in which empty-handed movements have the capacity to be interpreted in terms of representational goals. Humans spontaneously describe some empty-handed movement as representational, giving credence to gesture as a unique movement category. Our results also show that observers predominately interpret empty-handed movements as representational when objects are present, but not acted on. In addition, we replicate Schachner and Carey (2013), by showing that observers predominately attribute a movement-based goal to an actor who produces movements without any objects present, although some participants do describe movement in this context as representational.

Study 1 therefore establishes that empty-handed movements can be seen as representational, but also as movement for the sake of movement. In Study 2, we explore contextual cues that have the potential to make the empty-handed movements appear more meaningful. We predict that these cues will then encourage observers to interpret the empty-handed movements as representational.

3. Study 2

In Study 2, we explore the effect of context on the interpretation of empty-handed movements. Our framework predicts that the richer the context in which an empty-handed movement occurs, the more likely it is that the movement will be interpreted as representational, rather than as movement for its own sake. This hypothesis was supported by our results from Study 1—having objects in the scene provided a richer context that elicited more representational goal responses than having no objects present. Next, we test the effect of three contextual cues that should help viewers differentiate between interpreting movement in terms of representational goals versus movement-based goals. We aim to replicate our results on the presence or absence of objects, and consider two additional cues

—the form of the movement itself (in particular, the shape of the hand) and communicative intent (whether the movement is accompanied by speech).

Study 1 gave us reason to believe that the presence of objects can have a strong influence on whether empty-handed movements are seen as representational or movement-based. Performing an empty-handed movement in the presence of an object makes it easier to glean meaning from that movement than if it were performed without the object present. For example, if a twisting movement is performed near a jar, an observer is more likely to interpret the movement as a gesture for jar-opening than if it is performed without the jar present (the twister might just be flexing her fingers). Thus, as in Study 1, the first factor that we vary is the presence or absence of objects, none of which are ever touched. We hypothesize that observers will be more likely to interpret an empty-handed movement as a gesture if the movement is seen as meaningful; producing the movement near a relevant object is likely to increase its meaningfulness.

A second way that the empty-handed movements may be informative is in how closely the movement resembles actual movement needed to act on the objects—in other words, the precision or specificity of the movement. A rotating motion produced with a grasping handshape resembles the act of jar-opening, making it easier to glean meaning from the movement—easier than if the rotating movement were produced with an index finger handshape. Thus, the second factor we vary is the shape of the hand used in the movement. In particular, we vary whether the hand is shaped as it would have been had the movement been performed directly on the object—a grasping handshape vs. an index finger that traces the path, a tracer handshape. We hypothesize that observers will be more likely to interpret an empty-handed movement as a gesture if it is produced with a meaningful handshape (in this case, a grasping handshape) than if it is produced with a handshape that is more difficult to interpret (a tracer handshape).

Finally, most spontaneous gestures are produced along with speech (McNeill, 1992) and are thus part of a communicative act. Typically, the speech that accompanies a gesture is relevant to that gesture; the information conveyed in gesture either complements or supplements the information conveyed in speech (Goldin-Meadow, 2003). As we were interested in whether the presence of speech (rather than its content) affects when an empty-handed movement is interpreted as a gesture, we filtered speech to retain its prosody but render it unintelligible. We hypothesize that observers will be more likely to interpret an empty-handed movement as representational if it is seen as part of a communicative act, that is, if it is produced along with speech (even if the speech is uninterpretable), than if it is produced in silence.

3.1. Method

3.1.1. Participants—An additional 320 adult English speakers participated in Study 2 (40 in each of 8 conditions; 166 females, 150 males; 4 unreported), through the Amazon Mechanical Turk website (<https://www.mturk.com>). Requirements for participation and compensation were identical to Study 1. An additional 47 participants were excluded if they failed a prescreening audio check or had other technical difficulties viewing stimuli ($n = 5$), or if they had previously completed Study 1 ($n = 42$). As in Study 1, participants were over

18 years of age, and were ethnically diverse—White ($n = 224$); Black ($n = 38$), Asian ($n = 22$); Native American ($n = 4$); Native Pacific Islander ($n = 2$); more than one race ($n = 21$); unreported ($n = 9$). Participants took just over 4 min to complete the task.

3.1.2. Stimuli—There were eight videos used in Study 2 (see Fig. 3 for stills taken from the videos). Videos were structured as in Study 1, showing an actor performing the same movements viewed by participants in Study 1, and maintaining the timing of the movements across videos. The videos varied along 3 dimensions: (1) *Handshape* (Grasping handshape vs. Tracer handshape): The actor produced the movements using either a grasping handshape (i.e., the same palm-down C-hand shape used in Study 1) or a tracer hand shape (i.e., a pointing hand shape). (2) *Object Presence* (+Objects vs. –Objects): The actor produced the movements near, but not on, the same objects used in Study 1 (i.e., 4 balls, 2 boxes) or without any objects present. (3) *Speech Presence* (+Speech vs. –Speech): The actor produced the movements with speech or without it. The speech described what the actor was doing with her hands as she was moving them (see Study 1); the speech was filtered to retain prosody but be unintelligible. Filtering was achieved through the program Praat (a rectangular band filter was applied to frequencies less than 450 Hz), and individuals naïve to the original recording confirmed that the speech sounded muffled and was not comprehensible. The –Speech videos contained no sound.

3.1.3. Procedure—The procedure was identical to Study 1 with one exception: Participants viewing +Speech videos were given an audio file and asked to transcribe what they heard to confirm that the audio was functioning properly on their computers. They were permitted to listen to the audio file, which contained a string of 6 numbers, as many times as needed. Participants who failed to report the correct numbers were excluded from analyses ($n = 5$).

3.1.4. Coding and reliability—As in Study 1, participants' responses were coded as describing either (1) *External goals*, (2) *Representational goals*, (3) *Movement-based goals*, or (4) as *Uncodable*. Only 7 responses (2%) were uncodable. Again, two researchers who were blind to the experimental conditions of the participants coded each response, and each response was only assigned a single code, using the same method as Study 1. Coders agreed on 94% of trials, $K = .89$, and any coding disagreements were discussed and resolved by the researchers.

3.2. Results

Our goal was to determine the conditions under which observers identify a movement as a gesture. To that end, we tabulated the proportion of participants who described a *representational goal* in each of the 8 conditions (see Fig. 4). As would be expected from Study 1, almost no participants described *external goals* for the empty-handed movements in Study 2. When participants did not describe *representational goals*, they primarily described *movement-based goals*. We therefore conducted a 2 (Handshape: Grasping, Tracer) \times 2 (Speech Presence: +Speech, –Speech) \times 2 (Object Presence: +Objects, –Objects) factorial ANOVA with *representational goals* (0 or 1) as the dependent measure. The ANOVA revealed main effects of Handshape, $F(1, 318) = 45.514, p < .001$, and Object Presence, $F(1,$

318) = 17.645, $p < .001$, but no main effect of Speech Presence, $F(1, 318) = 0.739$, *ns*. As predicted, *representational goal* responses were higher when the actor used a grasping handshape than when she used a tracer handshape, and higher when objects were present than when they were absent. However, there was a significant interaction between Handshape and Speech Presence, $F(3, 316) = 8.571$, $p = .004$. No other 2- or 3-way interactions were significant.

To explore the interaction between Handshape and Speech Presence, we conducted two independent-sample t-tests examining the presence or absence of speech, first for all participants who viewed a Grasping Handshape, and second for all participants who viewed a Tracer Handshape. For participants who viewed an actor using a Grasping Handshape, the presence or absence of speech did not affect the proportion of participants who described *representational goals*, $t(158) = 1.27$, *ns*. Observers were likely to attribute representational goals to the agent when the movement was performed by a grasping handshape, whether or not speech was present. In contrast, for participants who viewed an actor using a Tracer Handshape, the presence or absence of speech did affect the proportion of participants who described *representational goals*, $t(158) = 2.99$, $p < .01$. Observers were likely to attribute representational goals to the agent when the movement was performed by a tracer handshape only when speech was present.

Together, our results confirm that when observers view movements that are not used to directly manipulate objects (empty-handed movements), multiple factors affect whether the movement is seen as a gesture. (1) The main effect of Object Presence replicates Study 1, and suggests that if empty-handed movements are performed in a meaningful context, in this case, in the presence of relevant objects, those movements are likely to be considered representations of acts. (2) The main effect of Handshape suggests that if empty-handed movements are performed with a meaningful handshape, one that resembles the handshape used when actually moving the objects, those movements are likely to be considered representations of acts. (3) The interaction between Handshape and Speech Presence suggests that if empty-handed movements are performed with a difficult-to-interpret handshape (in this case, a tracer handshape), those movements are more likely to be considered representations of acts if they are produced in the presence of speech, even if it is unintelligible. In contrast, when the handshape is meaningful (a grasping handshape), empty-handed movements tend to be considered representations of acts with or without the presence of speech.

4. Discussion

When do observers interpret a movement as representation of movement—as a gesture—rather than as a movement per se, produced either to affect an object or for its own sake? Our findings support a framework in which observers view empty-handed movements as having representational goals if those movements can be interpreted as meaningful. In Study 1, we replicated previous work (Schachner & Carey, 2013) showing that movement produced on an object is likely to be interpreted as a goal-directed action on that object, whereas movement produced in the absence of an object is likely to be interpreted as movement for the sake of movement. Importantly, we also tested participants'

interpretations of a third type of movement—movement produced off of an object (empty-handed movement), but in the presence of the object—and found that, under these circumstances, observers are likely to see the movement as a representation, that is, as a gesture for action, rather than as an action per se. Pushing the envelope further, in Study 2 we explored how contextual cues that make a movement appear more meaningful can work together to encourage observers to give representational responses for empty-handed movement. We found that observers are also likely to see an empty-handed movement as a gesture when an actor uses a hand-shape that resembles the handshape that would be used to manipulate the object as it is moved. If the actor uses a handshape that does not capture features of the hand as it manipulates the object (i.e., if it is a pointing index finger), the empty-handed movement will be interpreted as a gesture only if the movement is accompanied by speech-like sounds.

Our findings thus corroborate Schachner and Carey's (2013) claim that there are nuances to how movements are interpreted, and that observers do not inevitably see movement as having an external goal. The results further enrich the literature on action understanding by detailing a set of circumstances under which movement is likely to be interpreted as a representation of action (as gesture), rather than as action itself. The results also inform traditional view of gesture. Although gesture is often talked about as movements that accompany speech⁴ (Hostetter & Alibali, 2008; Kendon, 2004; McNeill, 1992), our results emphasize that speech is not always necessary for a movement to be seen as representational. Instead, contextual factors and top-down processes can lead to the categorization of movement as representational, and therefore gesture. Our findings can be used to create a more general framework for understanding how empty-handed movements are interpreted, and have implications for education and for how the differences between gesture and other forms of movement are conceptualized.

4.1. How context drives classification of empty-handed movement

Using the features in the present study as examples, we can consider a general framework for how different types of cues can drive representational versus movement-based goal interpretations of empty-handed movements. More specifically, in this framework, the richer the context, the more likely an empty-handed movement will be interpreted as meaningful, and thus as representational, rather than as movement for the sake of movement. When interpreting another individual's movements, observers can attend to contextual cues arising from a variety of sources. (1) *Cues internal to the movement*. First, cues to meaning can come from properties of the movement itself (e.g., the shape of the hand as it moves, the path of the movement, its rhythm or speed). If these properties resemble properties of a movement that might actually be performed on an object, the movement is more likely to be seen as meaningful, and thus representational. In the present study, handshape was this type of cue—handshape affords a way for an individual to interact with the external world and thus provides cues about that interaction. (2) *Cues external to the movement*. Second, cues to

⁴One type of gesture, called *emblems* (Ekman & Friesen, 1969, e.g., thumbs-up or the okay gesture), is often produced without speech; because emblems have standardized forms, they can be interpreted without speech. Our paper focuses on iconic gestures (*illustrators* in Ekman & Friesen's terms) that are tied to speech but, as we show here, can be recognized as a gesture even if processed without speech.

meaning can come from the environment within which a movement is produced. In the present study, the presence of objects was this type of cue. It is likely that an observer's prior knowledge of a particular object will affect her ability to use that object to interpret a movement. Adults know that balls can be picked up, and this knowledge likely drove their interpretation of the movements above the balls as representing actions to move these objects. If the agent had produced the same movements above novel objects (whose affordance was more difficult to guess from their appearance), observers might have been more inclined to attribute movement-based goals to the agent (e.g., she's waving her hands above some strange looking items). (3) *Communicative cues*. Third, cues to meaning can come from features that signal that the movement is part of a communicative act. In the present study, speech was this type of cue (although participants did not see the face of the person in the videos, the speech-like stimulus was interpreted as coming from her). Other communicative cues can include the specific content of the speech, the agent's facial expressions and eye-gaze, or even just the presence of a listener. Finally, we can consider how these different types of cues can interact. In the present study, cues to meaning for the tracer hand (an ambiguous handshape cue internal to the movement) were enhanced when the movement was produced along with speech-like sounds (a communicative cue). Thus, a combination of different types of cues can make a movement appear more meaningful and, as a result, lead an observer to interpret the movement as representational.

We have focused primarily on how various contextual cues can make it more likely for an observer to interpret a movement as representing an external goal. However, the framework can also be applied to interpreting a movement as representing a movement-based goal. For example, if an individual points left and then right, and then right again (a cue internal to the movement), and this is the only cue present, it is likely that the movements would be interpreted in terms of a movement-based goal. But if the same movements were performed along with music (a cue external to the movement) and/or with the speech, "So I have to move left, then right, then right again" (a communicative cue), the movements would likely be interpreted in terms of a representational goal, in this case, representing movement for its own sake (e.g., she was demonstrating steps in a dance, cf. Kirsh, 2010).

4.2. Implications for education

Over a decade of research has shown that gesture can facilitate learning (e.g., Alibali & DiRusso, 1999; Perry, Church, & Goldin-Meadow, 1988; Ping & Goldin-Meadow, 2010; Singer & Goldin-Meadow, 2005; Valenzeno, Alibali, & Klatzky, 2003; Wakefield & James, 2015), but the mechanisms underlying this powerful effect are still uncertain. One of the reasons gesture is thought to benefit learners is because it represents information about the to-be-learned concept in an easily accessible format (e.g., Goldin-Meadow, 2010). Gesture is assumed to represent concepts for the learner, as opposed to being mere hand-waving. But the learner may not always see it that way.

Our findings suggest that there are circumstances under which learners are more likely to see movement as gesture, and perhaps then profit from that movement. For example, in Studies 1 and 2, we saw that movements made near (but not directly on) objects were more likely to be interpreted as a representation than movements performed in the absence of

objects. Classroom teachers naturally gesture near objects (Alibali & Nathan, 2012), and most of the gesture strategies used in experimental situations to investigate the utility of teachers' gestures have also been performed in reference to objects (Cook, Mitchell, & Goldin-Meadow, 2008; Goldin-Meadow, Cook, & Mitchell, 2009; Goldin-Meadow et al., 2012; Novack et al., 2014; Valenzeno et al., 2003). Our results suggest that part of the reason gesture may have helped children in previous studies is because the proximity of the gesture to the object it was referencing encouraged children to interpret the movement as a meaningful gesture.

Given our findings, educators should be informed that proximity to referents does matter when using gesture as a teaching tool. If proximity is not possible, other methods of making it clear that a movement is representing information may be able to be used. For example, Ping and Goldin-Meadow (2008) showed that children can learn from gesture when objects are absent; however, before the children were shown the gestures, they had been familiarized with the objects about which the experimenter gestured, and thus could interpret the experimenter's movements in the context of those now-absent objects. Producing a movement in the presence of objects (or being familiar with those objects if they are not present) is likely to allow learners to glean meaning from that movement and thus interpret it as gesture. When a movement is produced in the absence of objects, learners may be more likely to see it as movement for its own sake, and thus not reap educational benefits from it.

A second cue that influenced whether a movement was interpreted as a representation was the handshape used by the actor in Study 2. Participants were significantly more likely to view an empty-handed movement as a representation if the actor used a grasping handshape during the movement than if she used a tracer handshape, whether or not objects were present. Although both grasping and tracer handshapes can render an empty-handed movement a representation (particularly if the tracer handshape is accompanied by speech-like sounds), the two handshapes are not interchangeable in the eyes of the participants. Paying attention to the form of an empty-handed movement (as well as its context of use) thus appears to be important in determining whether it will be interpreted as a gesture and therefore relevant to a lesson.

The participants in our study (and in Schachner & Carey's, 2013, study) were all adults. The cues that are important in rendering a movement a representation for an adult may not be relevant to children younger than 3 years, for whom representational thinking is challenging (e.g., Blades & Cooke, 1994; DeLoache, 1987; Simcock & DeLoache, 2006; Tolar, Lederberg, Gokhale, & Tomasello, 2008). For example, 2-year-olds with no previous Sign Language exposure perform at chance if shown a sign for an object in American Sign Language that is transparently related to the object, and asked to identify the object in a set of 4 items (Tolar et al., 2008). Young children's failure in studies of this sort is often explained by an underdeveloped symbolic mapping system (DeLoache, 1995), which raises the question of how young children would interpret the movements in our study. One possibility is that, even before children gain the capacity to interpret the symbolic form of a particular gesture, they will still identify an empty-handed movement as a representation of *something*. That is, there may be a property of movement (such as being performed near, but not on, an object) that encourages young children to categorize it as a representation, even if

they cannot decode the representation. Alternatively, young children may view empty-handed movement as movement for its own sake (cf., Schachner & Carey, 2013) until they gain more sophisticated representational abilities. As the majority of research on infants' and young children's interpretation of movement has been within the framework of external goals and actions on objects (e.g., Woodward, 1998), we know little about when young children interpret a movement as a representation (but see Novack, Goldin-Meadow, & Woodward, 2015). Although it is clear that young children exposed to a conventional sign language like ASL can learn that language as naturally, and following approximately the same timetable, as children exposed to a spoken language (Lillo-Martin, 1999; Newport & Meier, 1985), the as-yet-unanswered questions are—when do children interpret movements that are not part of a linguistic system as representations, and are the same cues to representation that we have found to be important to adults also important to children?

4.3. The relation between gesture and speech

The final factor we considered in Study 2 was the presence or absence of speech. We assume that if the speech that accompanies an empty-handed movement were to explicitly refer to that movement (e.g., "I'm moving the orange ball to the orange box"), the movement would be interpreted as a gesture. Our question was whether the presence of speech (rather than its content) would affect when an empty-handed movement is interpreted as a gesture. We therefore filtered the speech to retain its prosody but render it unintelligible. We predicted that the presence of these speech-like sounds would significantly increase the likelihood that an empty-handed movement would be interpreted as a gesture. However, we found this to be the case only when an ambiguous handshape was used in the movement. In other words, making an empty-handed movement seem more communicative in the context of speech-like sounds does not increase its likelihood of being considered gesture unless the movement is difficult to interpret in the first place.

We do know that the participants in our study saw the speech-like sounds as communicative—they thought that someone was trying to say something to them that they could not understand (e.g., "A woman's voice was garbled, like she was underwater"). The fact that this communicative, albeit unintelligible, cue did not generally increase the likelihood that an empty-handed movement would be interpreted as a gesture (only when the gesture was ambiguous) suggests that the aspect of speech that drives individuals to interpret movement as meaningful gesture is not just the intent to communicate, but also the meaning of the spoken language itself. Neuroimaging results lend weight to this interpretation—using fMRI, researchers have found a different pattern of activation when participants process empty-handed movements that are produced with a spoken language they know, compared to the same movements produced with a language they do not know (Green et al., 2009). This finding has interesting implications for using gesture to teach a second language. We know that producing movements along with the sounds of a new language can help individuals learn that language (Kelly, McDevitt, & Esch, 2009; Macedonia, Muller, & Friederici, 2011), but it may be important to tell the learners that the movements are meaningful (i.e., are gestures) in order for them to be effective.

5. Conclusions

Research on how we understand action across the lifespan has a long and rich history, but until recently was narrow in scope, focusing predominately on actions used to manipulate objects. Our study is only the first step in exploring the features of movement that make it likely to be interpreted as a representation. This work expands the discussion of how humans accomplish the task of interpreting the actions of others by investigating actions that are not used to manipulate objects. We show that these understudied, empty-handed actions have the capacity to be interpreted either as movement for the sake of movement (Schachner & Carey, 2013) or as a representation. This distinction is not trivial—when movements are seen as representations, they have the power to influence communication, learning, and cognition in ways that movement for its own sake does not. By incorporating gesture into a framework for movement analysis, we take an important step towards developing a more cohesive understanding of action-interpretation.

References

- Alibali MW, DiRusso AA. The function of gesture in learning to count: More than keeping track. *Cognitive Development*. 1999; 14:37–56. [http://dx.doi.org/10.1016/S0885-2014\(99\)80017-3](http://dx.doi.org/10.1016/S0885-2014(99)80017-3).
- Alibali MW, Nathan MJ. Embodiment in mathematics teaching and learning: Evidence from learners' and teachers' gestures. *Journal of the Learning Sciences*. 2012; 21:247–286. <http://dx.doi.org/10.1080/10508406.2011.611446>.
- Baldwin DA, Baird JA. Discerning intentions in dynamic human action. *Trends in Cognitive Sciences*. 2001; 5:171–178. [http://dx.doi.org/10.1016/S1364-6613\(00\)01615-6](http://dx.doi.org/10.1016/S1364-6613(00)01615-6). [PubMed: 11287271]
- Baldwin DA, Baird JA, Saylor MM, Clark MA. Infants parse dynamic action. *Child Development*. 2001; 72:708–717. <http://dx.doi.org/10.1111/1467-8624.00310>. [PubMed: 11405577]
- Blades M, Cooke Z. Young children's ability to understand a model as a spatial representation. *The Journal of Genetic Psychology: Research and Theory on Human Development*. 1994; 155:201–218. <http://dx.doi.org/10.1080/00221325.1994.9914772>.
- Bower GH, Rinck M, Goldman SR, Graesser AC, van den Broek P. Goals as generators of activation in narrative understanding. *Narrative comprehension, causality, and coherence: Essays in honor of Tom Trabasso*. 1999:111–134. Erlbaum.
- Buresh, JS.; Woodward, A. Infants track action goals within and across agents; *Cognition*. 2007. p. 287-314. <http://dx.doi.org/10.1016/j.cognition.2006.07.001>
- Church RB, Kelly SD, Wakefield EM, Matsumoto D, Hwang HC, Frank MG. *Measuring gesture*. APA handbook of non-verbal communication. 2015
- Cook SW, Mitchell Z, Goldin-Meadow S. Gesturing makes learning last. *Cognition*. 2008; 106:1047–1058. <http://dx.doi.org/10.1016/j.cognition.2007.04.010>. [PubMed: 17560971]
- Crump MJ, McDonnell JV, Gureckis TM. Evaluating Amazon's Mechanical Turk as a tool for experimental behavioral research. *PLoS One*. 2013; 8:e57410. <http://dx.doi.org/10.1371/journal.pone.0057410>. [PubMed: 23516406]
- DeLoache J. Rapid change in the symbolic functioning of very young children. *Science*. 1987; 238:1556–1557. <http://dx.doi.org/10.1126/science.2446392>. [PubMed: 2446392]
- DeLoache J. Understanding and use of symbols: The model model. *Current Directions in Psychological Science*. 1995; 4:109–113. <http://dx.doi.org/10.1111/1467-8721.ep10772408>.
- Dick AS, Goldin-Meadow S, Hasson U, Skipper JI, Small SL. Co-speech gestures influence neural activity in brain regions associated with processing semantic information. *Human Brain Mapping*. 2009; 30:3509–3526. <http://dx.doi.org/10.1002/hbm.20774>. [PubMed: 19384890]
- Ekmann P, Friesen W. The repertoire of nonverbal behavior: Categories, origins, usage, and coding. *Semiotica*. 1969; 1:49–98. <http://dx.doi.org/10.1515/semi.1969.1.1.49>.

- Goldin-Meadow, S. *Hearing gesture: How our hands help us think*. Belknap Press of Harvard University Press; Cambridge, M.A.: 2003.
- Goldin-Meadow S. When gesture does and does not promote learning. *Language Cognition*. 2010; 2:1–19. <http://dx.doi.org/10.1515/LANGCOG.2010.001>. [PubMed: 20640053]
- Goldin-Meadow S, Cook SW, Mitchell Z. Gestures gives children new ideas about math. *Psychological Science*. 2009; 20:267–271. <http://dx.doi.org/10.1111/j.1467-9280.2009.02297.x>. [PubMed: 19222810]
- Goldin-Meadow S, Levine SC, Zinchenko E, Yip TK, Hemani N, Factor L. Doing gesture promotes learning a mental transformation task better than seeing gesture. *Developmental Science*. 2012; 15:876–884. <http://dx.doi.org/10.1111/j.1467-7687.2012.01185.x>. [PubMed: 23106741]
- Goldin-Meadow S, Sandhofer CM. Gestures convey substantive information about a child's thoughts to ordinary listeners. *Developmental Science*. 1999; 2:67–74. <http://dx.doi.org/10.1111/1467-7687.00056>.
- Green A, Straube B, Weis S, Jansen A, Willmes K, Konrad K, Kircher T. Neural integration of iconic and unrelated coverbal gestures: A functional MRI study. *Human Brain Mapping*. 2009; 30:3309–3324. <http://dx.doi.org/10.1002/hbm.20753>. [PubMed: 19350562]
- Hostetter AB, Alibali MW. Visible embodiment: Gestures as simulated action. *Psychonomic Bulletin & Review*. 2008; 15:495–514. <http://dx.doi.org/10.3758/pbr.15.3.495>. [PubMed: 18567247]
- Kelly, SD.; Healy, M.; Ozyurek, A.; Holler, J. The processing of speech, gesture, and action during language comprehension. *Psychonomic Bulletin & Review*. 2014. <http://dx.doi.org/10.3758/s13423-014-0681-7>
- Kelly SD, McDevitt T, Esch M. Brief training with co-speech gesture lends a hand to word learning in a foreign language. *Language and Cognitive Processes*. 2009; 24:313–334. <http://dx.doi.org/10.1080/01690960802365567>.
- Kendon A. Do gestures communicate? A review. *Research on Language and Social Interaction*. 1994; 27:175–200. http://dx.doi.org/10.1207/s15327973rlsi2703_2.
- Kendon, A. *Gesture: Visible action as utterance*. The University of Chicago Press; Chicago, IL: 2004.
- Kirsh, D. Thinking with the body. In: Ohlsson, S.; Catrambone, R., editors. *Proceedings of the 32nd annual conference of the cognitive science society*. Cognitive Science Society; Austin, TX: 2010. p. 2864-2869.
- Lillo-Martin, D. Modality effects and modularity in language acquisition: The acquisition of American Sign Language. In: Ritchie, WC.; Bhatia, TK., editors. *Handbook of language acquisition*. Academic Press; San Diego, C.A.: 1999. p. 531-567.
- Macedonia M, Muller K, Friederici AD. The impact of iconic gestures on foreign language word learning and its neural substrate. *Human Brain Mapping*. 2011; 32:982–998. <http://dx.doi.org/10.1002/hbm.21084>. [PubMed: 20645312]
- McNeill, D. *Hand and mind: What gestures reveal about thought*. The University of Chicago Press; Chicago, IL: 1992.
- Newport, EL.; Meier, RP. The acquisition of American Sign Language. In: Slobine, DI., editor. *The cross-linguistic study of language acquisition*. Vol. 1. Lawrence Erlbaum Associates; Hillsdale, N.J.: 1985. p. 881-938.
- Newton, D. Foundations of attribution: The perception of ongoing behavior. In: Harvey, JH.; Ickes, WJ.; Kidd, RF., editors. *New directions in attribution research*. Erlbaum; Hillsdale, N.J.: 1976. p. 223-248.
- Novack, M.; Congdon, E.; Hemani-Lopez, N.; Goldin-Meadow, S. From action to abstraction: Using the hands to learn math. *Psychological Science*, online first. 2014. <http://dx.doi.org/10.1177/0956797613518351>
- Novack M, Goldin-Meadow S, Woodward A. Learning from gesture: How early does it happen? *Cognition*. 2015; 142:138–147. <http://dx.doi.org/10.1016/j.cognition.2015.05.018>. [PubMed: 26036925]
- Perry M, Church RB, Goldin-Meadow S. Transitional knowledge in the acquisition of concepts. *Cognitive Development*. 1988; 3:359–400. [http://dx.doi.org/10.1016/0885-2014\(88\)90021-4](http://dx.doi.org/10.1016/0885-2014(88)90021-4).

- Ping RM, Goldin-Meadow S. Hands in the air: Using ungrounded iconic gestures to teach children conservation of quantity. *Developmental Psychology*. 2008; 44:1277–1287. <http://dx.doi.org/10.1037/0012-1649.44.5.1277>. [PubMed: 18793062]
- Ping R, Goldin-Meadow S. Gesturing saves cognitive resources when talking about nonpresent objects. *Cognitive Science*. 2010; 34:602–619. <http://dx.doi.org/10.1111/j.1551-6709.2010.01102.x>. [PubMed: 21564226]
- Schachner A, Carey S. Reasoning about ‘irrational’ actions: When intentional movements cannot be explained, the movements themselves are seen as the goal. *Cognition*. 2013; 129:309–327. <http://dx.doi.org/10.1016/j.cognition.2013.07.006>. [PubMed: 23958932]
- Searle JR. The intentionality of intention and action. *Cognitive Science*. 1980; 4:47–70. http://dx.doi.org/10.1207/s15516709cog0401_3.
- Simcock G, DeLoache J. Get the picture? The effects of iconicity on toddlers’ reenactment from picture books. *Developmental Psychology*. 2006; 42:1352–1357. <http://dx.doi.org/10.1037/0012-1649.42.6.1352>. [PubMed: 17087568]
- Singer MA, Goldin-Meadow S. Children learn when their teacher’s gestures and speech differ. *Psychological Science*. 2005; 16:85–89. <http://dx.doi.org/10.1111/j.0956-7976.2005.00786.x>. [PubMed: 15686572]
- Tolar TD, Lederberg AR, Gokhale S, Tomasello M. The development of the ability to recognize the meaning of iconic signs. *Deaf Studies and Deaf Education*. 2008; 13:225–240. <http://dx.doi.org/10.1093/deafed/enm045>. [PubMed: 17827446]
- Trabasso T, Nickels M. The development of goal plans of action in the narration of a picture story. *Discourse Processes*. 1992; 15:249–275. <http://dx.doi.org/10.1080/01638539209544812>.
- Trabasso T, Stein NL, Rodkin PC, Munger MP, Baughn CR. Knowledge of goals and plans in the on-line narration of events. *Cognitive Development*. 1992; 7:133–170. [http://dx.doi.org/10.1016/0885-2014\(92\)90009-G](http://dx.doi.org/10.1016/0885-2014(92)90009-G).
- Trofatter, C.; Kontra, C.; Beilock, SL.; Goldin-Meadow, S. Gesture has a larger impact on problem-solving than action, even when action is accompanied by words. *Language, Cognition and Neuroscience*. 2014. <http://dx.doi.org/10.1080/23273798.2014.905692>
- Valenzeno L, Alibali MW, Klatzky R. Teachers’ gestures facilitate students’ learning: A lesson in symmetry. *Contemporary Educational Psychology*. 2003; 28:187–204. [http://dx.doi.org/10.1016/s0361-476x\(02\)00007-3](http://dx.doi.org/10.1016/s0361-476x(02)00007-3).
- Wakefield EM, James KH. Effects of learning with gesture on children’s understanding of a new language concept. *Developmental Psychology*. 2015; 51:1105–1114. <http://dx.doi.org/10.1037/a0039471>. [PubMed: 26214229]
- Woodward A. Infants selectively encode the goal object of an actor’s reach. *Cognition*. 1998; 69:1–34. [http://dx.doi.org/10.1016/S0010-0277\(98\)00058-4](http://dx.doi.org/10.1016/S0010-0277(98)00058-4). [PubMed: 9871370]
- Zacks J, Tversky B, Iyer G. Perceiving, remembering and communicating structure in events. *Journal of Experimental Psychology: General*. 2001; 130:29–58. <http://dx.doi.org/10.1037/0096-3445.130.1.29>. [PubMed: 11293458]

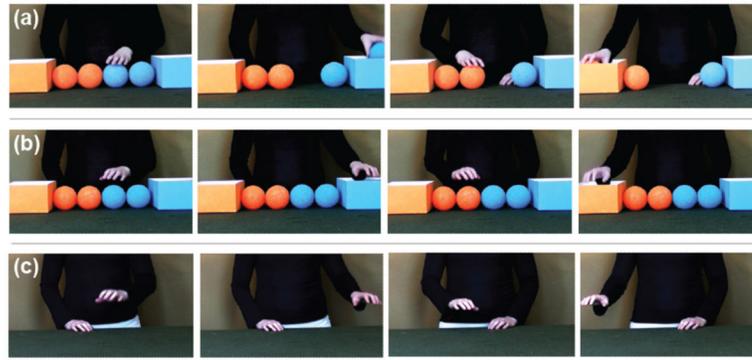


Fig. 1. Stills from video stimuli in the three conditions in Study 1: (a) Action on Objects, (b) Action off Objects with Objects Present, (c) Action with Objects Absent.

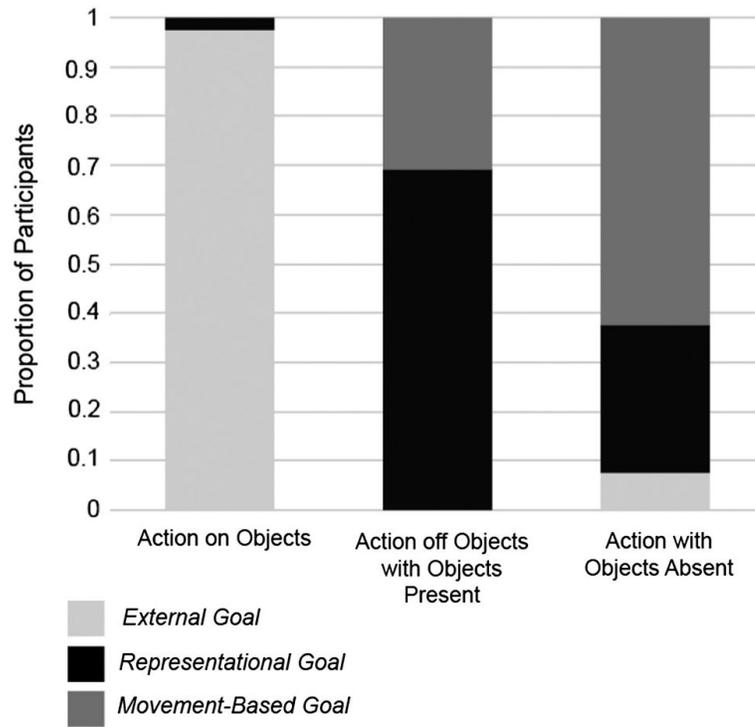


Fig. 2. Proportion of participants in each condition giving External, Representational or Movement-based goal responses.

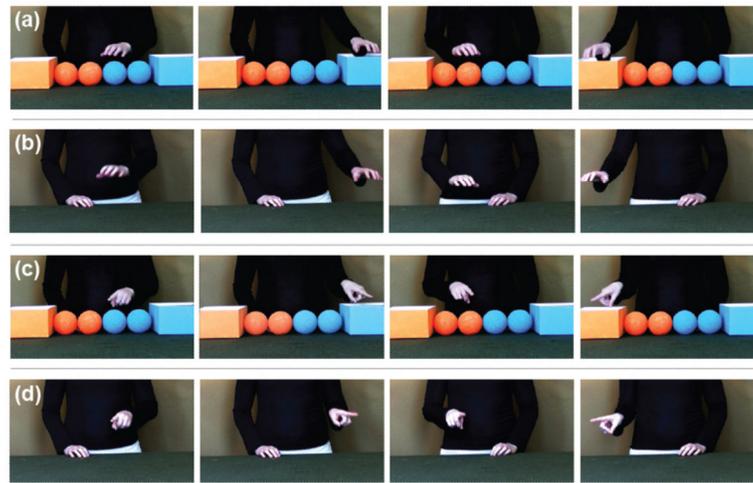


Fig. 3. Stills from the video stimuli in Study 2: (a) Grasping Handshape +Objects, (b) Grasping Handshape –Objects, (c) Tracer Handshape +Objects, (d) Tracer Handshape –Objects. Two version of each video were created (+Speech, –Speech), resulting in 8 videos.

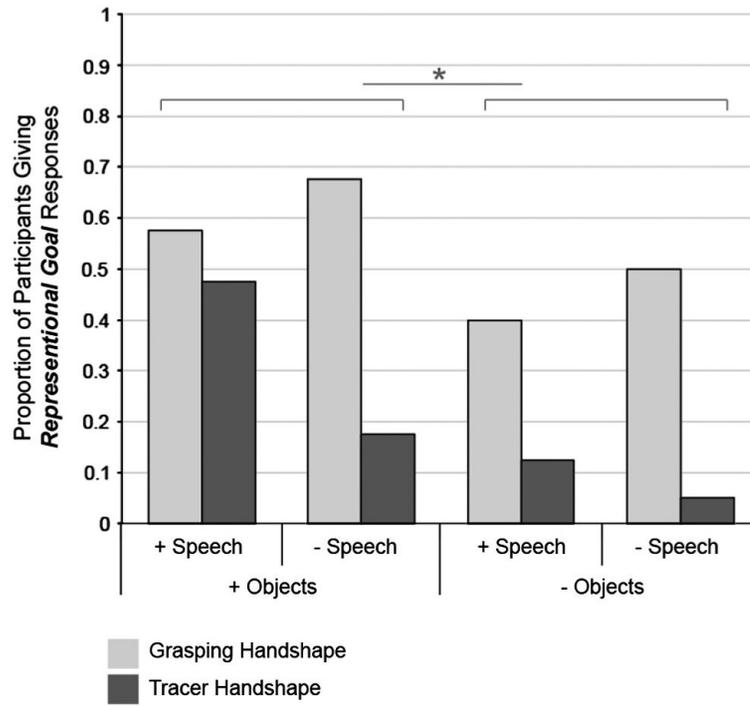


Fig. 4. Proportion of participants giving Representational goal responses in each of the 8 conditions.