





# When gestures show us the way: Co-thought gestures selectively facilitate navigation and spatial memory

Alexia Galati <sup>a,b</sup>, Steven M. Weisberg<sup>c</sup>, Nora S. Newcombe <sup>d</sup>,  
and Marios N. Avraamides<sup>a,e</sup>

<sup>a</sup>Department of Psychology, University of Cyprus, Nicosia, Cyprus; <sup>b</sup>Cognitive and Information Sciences, University of California–Merced, Merced, CA, USA; <sup>c</sup>Department of Neurology, University of Pennsylvania, Philadelphia, PA, USA; <sup>d</sup>Department of Psychology, Temple University, Philadelphia, PA, USA; <sup>e</sup>Centre for Applied Neuroscience, University of Cyprus, Nicosia, Cyprus

## ABSTRACT

How does gesturing during route learning relate to subsequent spatial performance? We examined the relationship between gestures produced spontaneously while studying route directions and spatial representations of the navigated environment. Participants studied route directions, then navigated those routes from memory in a virtual environment, and finally had their memory of the environment assessed. We found that, for navigators with low spatial perspective-taking performance on the Spatial Orientation Test, more gesturing from a survey perspective predicted more accurate memory following navigation. Thus, co-thought gestures accompanying route learning relate to performance selectively, depending on the gesturers' spatial ability and the perspective of their gestures. Survey gestures may help some individuals visualize an overall route that they can retain in memory.

## KEYWORDS

gestures; navigation; spatial memory; route learning; individual differences

## 1. Introduction

People often have to follow route directions to navigate in an unfamiliar environment, whether these directions are provided by a friend over the phone, are announced piecemeal by a GPS system, or are printed on a party invitation. When navigators receive linguistic route directions in advance, they have the opportunities to study them to create a representation of the described route, which they can later retrieve from memory to navigate in the environment. But transforming linguistic information into spatial directions is difficult and potentially inexact, as when a right turn is a “right veer” and not a 90-degree turn. Moreover, while some people may prefer to re-code verbal information into a spatial form—by drawing, gesturing, or mentally simulating the described spatial relationships or movement—others may prefer to retain the steps as linguistic propositions. Here, we investigate the

**CONTACT** Alexia Galati  [agalati@ucmerced.edu](mailto:agalati@ucmerced.edu)  Cognitive & Information Sciences, SSM Building, University of California, Merced, 5200 N. Lake Rd., Merced CA 95343, USA.

Color versions of one or more of the figures in the article can be found online at [www.tandfonline.com/hssc](http://www.tandfonline.com/hssc).

© 2017 Taylor & Francis

role of gesture as a potential scaffold between language and internal representations of space, and whether its role may vary for different people. Specifically, we examine whether gesturing while learning verbal route directions of an unfamiliar environment is associated with better subsequent navigation performance and memory for that environment.

Gestures systematically accompany language that expresses spatial information (Alibali, 2005), including direction giving (Allen, 2003), and descriptions of spatial patterns (Melinger & Kita, 2007), shapes (Graham & Argyle, 1975), and motion in space (Kita & Özyürek, 2003). This tight link between gesturing and reasoning about space can be conceptualized within current models of gesture production. In some accounts, representational gestures—those depicting semantic content by virtue of their handshape, placement, and movement—facilitate performance by activating task-relevant features through mental (and neural) simulation of prior spatiomotor experiences (Hostetter & Alibali, 2008, 2010). In other accounts, gesturing relies on spatial imagery (de Ruiter, 2000; Kita & Özyürek, 2003; Krauss, Chen, & Gottesman, 2000) or recruits visuospatial working memory (Goldin-Meadow, 2011; Wesp, Hesse, Keutmann, & Wheaton, 2001).

Even in the absence of accompanying speech, self-generated gestures produced without communicative intent, often referred to as *co-thought gestures*, can be helpful in a host of situations that require spatial transformations. These include mental rotation (Chu & Kita, 2011), making spatiomotor inferences (e.g., about gear movements, Alibali, Spencer, Knox, & Kita, 2011) or making inferences about spatial relations from descriptions of environments (Jamalian, Giardino, & Tversky, 2013). In light of these findings, co-thought gestures in preparation for navigation might be expected to improve the spatial representation of the to-be-navigated route. One recent study by So and colleagues (So, Ching, Lim, Cheng, & Ip, 2014) offers some initial support for this proposal. In that study, after studying paths presented in diagrams, participants rehearsed the path by gesturing, drawing, or mentally simulating it with their hands held still. Participants recalled more steps of the path when they gestured during rehearsal than all other conditions, including drawing. The authors proposed that gesturing leads to deeper encoding due to its property of leaving no visible trace, which requires people to maintain an active internal image of the route sequence, whereas drawing externalizes the route sequence on paper and thus does not place such demands.

However, not all gesturers are alike. Individual differences in spatial and verbal abilities are associated with differences in gesture production (Alibali, 2005). Individuals' frequency of gesturing appears to be related to a combination of their spatial and verbal abilities: individuals with high spatial visualization ability but low phonemic fluency (the ability to organize ideas into a chain of linguistic units, associated with executive control) are more likely to gesture (Hostetter & Alibali, 2007).

Working memory capacity has also been linked to gesture production, with evidence for both verbal and visuospatial aspects of it predicting gesture. Some studies show that more co-speech gestures are produced by those with low verbal working memory capacity (Gillespie, James, Federmeir, & Watson, 2014), whereas other studies show that more representational gestures are produced by those with lower visual and spatial working memory (Chu, Meyer, Foulkes, & Kita, 2014).

Some recent work suggests that, at least for some tasks, there is an interaction between individual ability and co-speech gestures. Extending previous findings that gestures help performance in dual tasks (e.g., Goldin-Meadow, Nusbaum, Kelly, & Wagner, 2001), Marstaller and Burianova (2013) demonstrated that the effect of gesturing is related to individual differences in working memory capacity. When asked to recall a series of letters, individuals with low working memory capacity benefited from being able to gesture during an intervening explanation of how they solved a mathematical equation, whereas high-capacity individuals did not benefit from gesturing. In fact, low and high capacity individuals did not differ in their recall of letters when they could gesture while explaining their solution. These and other studies (e.g., Sassenberg, Foth, Wartenburger, & van der Meer, 2011) suggest that individual differences in spatial and other abilities can clarify the influence of gesturing on learning and performance.

In a previous experiment (Galati, Weisberg, Newcombe, & Avraamides, 2015), we investigated whether gesturing conferred a benefit to navigating an unfamiliar environment, depending on the navigators' individual abilities. Participants studied directions describing routes from a start point to a destination, then navigated those routes from memory in a virtual environment, and finally performed two memory tests that assessed their memory of the environment. In that study, we specifically instructed participants to use gesture while learning verbal directions for one route, and not to use gesture during direction learning for a second route.

In the gesture condition, participants were instructed to produce at least one compatible gesture for each numbered step of the route directions. Participants were shown examples of the types of gestures they might make (e.g., moving a sideways-oriented palm to carve a left turn; or drawing a left turn with an index finger on the desk). In the no-gesture condition, participants were required to keep their fingers on the keyboard, and were thus prevented from gesturing. At the beginning of the experiment, participants completed self-report and psychometric measures intended to capture individual differences in spatial ability, including three self-report measures—the Santa Barbara Sense of Direction test (SBSOD; Hegarty, Richardson, Montello, Lovelace, & Subbiah, 2002), the Philadelphia Spatial Ability Scale (PSAS; Hegarty, Crookes, Dara-Abrams, & Shipley, 2010), and the Philadelphia Verbal Ability Scale (PVA; Hegarty et al., 2010)—and one objective spatial test that assesses the ability to adopt imagined

spatial perspectives, the Spatial Orientation Test (SOT, Hegarty & Waller, 2004). We did not find a general benefit of forced gesturing (or general decrement of gesture restriction) in that study. Instead, we found evidence suggesting that individual differences interacted with gesturing to determine participants' final memory performance: at least for one of the two routes used, gesturing led to better memory, particularly for navigators with lower spatial ability scores.

These findings fit with the hypothesis that gesture may be beneficial for some but not for others. Allowing participants to gesture spontaneously could provide insight into the relationship between gesturing, navigation, and spatial memory under more naturalistic circumstances. In the present study, we used the same general method as Galati et al. (2015), but allowed participants to gesture freely as they studied both routes, without explicit instructions about gesturing. Participants could now self-select whether they would gesture (perhaps if they felt gesturing might benefit them) or not (if they felt gesturing incurred a dual-task cost).

We first aimed to establish whether, in the domain of route learning, spatial ability—and in particular spatial perspective-taking ability, as captured by the SOT—predicts how much people gesture and the perspective from which they do so. We were specifically interested in the relationship between spatial perspective-taking ability and the use of representational gestures involving a route or a survey perspective (Emmorey, Tversky, & Taylor, 2000), because these gestures could reflect representational strategies associated with differences in subsequent performance. When adopting a *route* perspective, the individual takes a mental tour of the environment from the navigator's perspective. Gestures instantiating this perspective typically capture the directional turns of the navigator, as with the sideways-oriented palm representing the left turn described above. When adopting a *survey* perspective, the imagined viewpoint is stationary and external to the environment. Gestures instantiating this perspective capture the path taken by the navigator from an external viewpoint, as when using an index finger to trace of the navigator's path on the table.

Insofar as gestures that embody a different spatial perspective reinforce different conceptualizations of spatial relationships (Emmorey, Tversky, & Taylor, 2000), we hypothesized that these conceptualizations, in turn, may be associated with differences in subsequent performance. In the context of learning route directions, route gestures might be effective for reinforcing the sequence of directional turns that learners need to keep track of for successful navigation, as these gestures typically express serialized bits of directional information (e.g., some sequence of left turns, right turns, and straight segments).

Although sequential route gestures depend on one another insofar as they require successfully updating the heading of the navigator at each step, the spatial information they convey is limited to the directionality of movement and is conveyed in a piecemeal fashion. To represent relationships between directional turns and other features of the environment, including landmarks, learners have to use survey gestures or other externalizing means, or else do so mentally. Survey gestures, although still evanescent, are tethered to a relatively stable gesture space, which may enable learners to outline the relationship between segments of the route and, by consequence, to better appraise relationships between landmarks encountered along the way. Through the use of a stable 2D gesture space, survey gestures can add spatial information to an externalized mental model of the environment that captures global relationships of features of the environment (e.g., interrelationships of landmarks or the overall path of the route).

In other contexts, learning an environment from a survey perspective (e.g., by studying a map) has been shown to provide a more complete representation of the environment than learning it from a route perspective (e.g., experiencing it through navigation) (Pazzaglia & Taylor, 2007). We therefore expected that creating a map-like representation with one's hands might have similarly beneficial effects, whether across the board or for specific groups of individuals (e.g., those with poor spatial perspective-taking ability).

We expected that, when people study route directions, their spatial perspective-taking ability might predict the gestures they produced and how much they benefit from gestures of a particular perspective. We therefore hypothesized that individual ability would moderate the relationship between gesturing and navigating. Additionally, we reasoned that, while navigating in an unfamiliar environment, navigators would update the initial representation they had constructed at study by using the visual information available in the environment. We therefore hypothesized that navigation performance would mediate the relationship between gesturing and memory performance. To examine these predictions, we assessed models that captured this overarching relationship between spatial perspective-taking ability, gesturing, navigation performance, and memory performance, and explored this relationship in separate models for gestures from a route and survey perspective.

## 2. Method

### 2.1. Participants

Twenty-four undergraduate and graduate students (12 female) from the University of Cyprus participated for research credit for a university course or for payment (15 euros).

## 2.2. Materials

### 2.2.1. Virtual environment

The VE was run on a Windows 7 processor [64-bit with an Intel Core i7 960 @ 3.20 GHz] with a [NVIDIA GeForce GTX 460] graphics card, and was projected on a 295 cm × 180 cm projection screen. The viewing distance from the projection screen was approximately 200 cm.

The VE (Virtual SILCton) was modeled after a real-world college campus and was created in Unity3D ([www.unity3d.com](http://www.unity3d.com)) using freely available buildings and objects from Google Sketchup (<http://sketchup.google.com/>) (for more information see Schinazi, Nardi, Newcombe, Shipley, & Epstein, 2013 and Weisberg, Schinazi, Newcombe, Shipley, & Epstein, 2014). Landmark buildings within were marked with a blue diamond and a nearby sign indicating the building's name.

### 2.2.2. Route descriptions

Routes were described as a series of steps (see Appendix A for an example). Because routes connected buildings in a preexisting VE, route descriptions could not be fully equivalent. As shown in Figure 1, Route 1 was slightly more complex than Route 2, involving 7 described turns (vs. 5 for Route 2) and 12 distinct segments of text in the directions (vs. 10 for Route 2). Nevertheless, the routes were matched in other dimensions: both connected four landmark buildings (including the origin and destination buildings), had 8 spatial locatives (e.g., *left*, *right*, *straight*) in their descriptions, and included 2 landmark buildings that were intervisible.



**Figure 1.** Aerial view map of the layout of buildings with two routes (solid lines) in the virtual environment. Route 1 is shown in red and Route 2 in blue.

## **2.3. Self-report and psychometric measures**

### **2.3.1. Santa Barbara sense of direction scale (SBSOD-CY)**

The SBSOD (Hegarty et al., 2002) is designed to measure participants' assessment of their own navigation abilities, with lower scores indicating lower navigation ability. The SBSOD consists of 15 items, involving some aspect of environmental spatial cognition (e.g., "I very easily get lost in a new city," "I am very good at judging distances"), to which participants respond on a 7-point Likert scale. Although participants in this study completed all 15 items of the SBSOD scale, subsequent analyses were based on that subset of items (referred to as the SBSOD-CY scale). This was because earlier work has suggested that only items 10 of the 15 items of SBSOD are suitable for measuring SOD in the Greek-Cypriot population (Shimi & Avraamides 2008).

### **2.3.2. Philadelphia spatial ability scale (PSAS)**

The PSAS scale (Hegarty et al., 2010) is designed to measure how well participants feel they can perform small-scale spatial tasks, such as visualizing and transforming small or medium-sized objects. It consists of 16 items (e.g., "I can easily visualize my room with a different furniture arrangement" and "I would be very good at building a model airplane, car, or train.") to which participants respond on a 7-point Likert scale.

### **2.3.3. Philadelphia verbal ability scale (PVA)**

The PVAS scale (Hegarty et al., 2010) is designed to measure how strong participants feel their verbal ability is. It consists of 10 items (e.g., "I am good at crossword puzzles" and "I would rather read a text explanation than look at a drawing or figure") to which participants respond on a 7-point Likert scale.

The SBSOD, PSAS, and PVA were translated into Greek and presented on a browser, using SurveyMonkey Inc. services.

### **2.3.4. Spatial Orientation Test (SOT)**

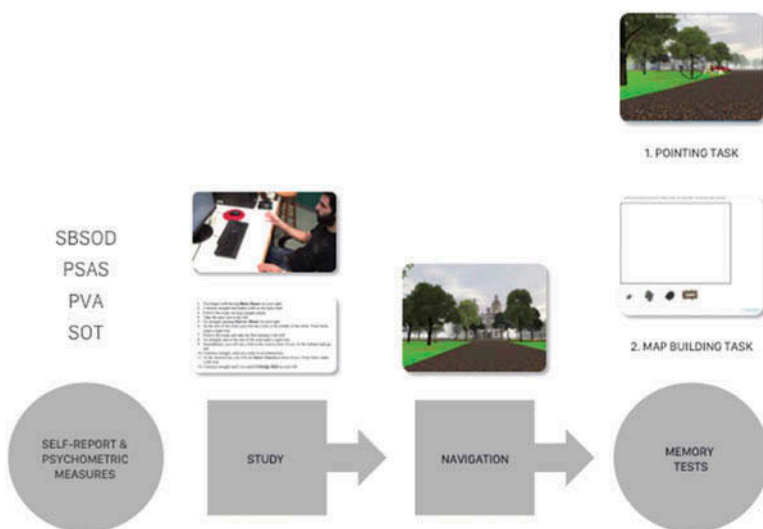
In the SOT (Hegarty & Waller, 2004), participants view an array of objects on a piece of paper and, on a given test item, they are asked to locate an object from an imagined perspective. To locate the third object, participants have to draw its angle of disparity from their imagined perspective on a circle on the page. Participants were timed for 5 minutes to complete as many of the 12 test items as they could. For each participant, the difference between the angle for the correct answer and their response was computed for each item, and was averaged across items to yield an overall error score.

## 2.4. Procedure

Participants were told that the study investigated navigation performance and were informed about the overall structure of the experiment. Upon giving informed consent, participants completed the self-report and psychometric measures (SBSOD, PSAS, PVA, and SOT). Next, participants were familiarized with the controls for moving in and looking around in the VE; the mouse was used to look and the arrow keys on a numeric keypad were used to move. Participants moved around the VE (in a section of it that was never encountered during the routes) until they felt comfortable with using the navigation controls. After that familiarization phase, participants first learned about the to-be-navigated environment by studying a verbal description of a route. Then, they navigated that route from memory in the VE, and finally completed two tests assessing their memory representation for the virtual environment. The main phases of the procedure are illustrated schematically in Figure 2 and are described in more detail next.

### 2.4.1. Study phase

After completing the questionnaires and psychometric measures and becoming familiarized with the VE controls, participants moved to an adjacent room to study the directions for one of the routes. This was done to ensure that participants used an enduring off-line spatial representation when



**Figure 2.** Outline of experimental procedure. Participants first completed the self-report and psychometric measures listed, and then, after a brief familiarization with the controls for moving in the VE (not shown), they studied the route directions (Study phase). Next, they navigated the route from memory in the VE (Navigation phase), and finally completed two tests assessing their memory performance for the environment (Memory tests phase). These three phases (Study, Navigation, Memory tests) were repeated for a second route.



navigating, as opposed to a transient sensorimotor representation (see Avraamides & Kelly, 2008, for a discussion). Participants were told that they would learn about a route through verbal descriptions and that they would later navigate the described path from memory in the virtual environment. They were informed that the route directions would be in Greek, but the four landmark buildings of the route would have English names; they were asked to ensure that they remembered those names in order to recognize the landmark buildings when encountering them in the VE. Participants were told that the route description would appear on the computer screen as a series of numbered instructions that would be “similar to format of directions from Google Maps, but more detailed.” Gesturing was not mentioned in the instructions.

After instructions, the experimenter turned on a digital camcorder (SONY HDR-CX 155) with a view of the participant and left the room. Participants pressed the spacebar to have the route directions appear on the computer screen and studied them without a time limit.

#### **2.4.2. Navigation phase**

After studying the route directions, participants returned to the original room to navigate in the VE. The viewer’s position and heading in the VE was already set at the origin building for that route. The experimenter reminded participants of where they were and where they needed to go (e.g., “You are at Batty House and you want to go to Golledge Hall.”), and explained that if they got lost they would be verbally directed to the last location in the VE where they had been correctly and be prompted to continue from there.

If during navigation participants made an error (e.g., taking a wrong turn) and did not readily self-correct, they were interrupted by the experimenter, were directed to an earlier point of the route (typically the point where the navigator was, after the previous instruction had been followed correctly), and were prompted for the next instruction. Prompts started as broad as possible (e.g., “you are now back at the intersection, do you remember what comes next?”), and became more specific only if the participant reported not being able to remember how to continue (e.g., “the next instruction says something about a museum. Do you remember now?”).

The participants’ navigation performance was video recorded by screen capture software, while the navigation session was also audiotaped.

#### **2.4.3. Memory tests**

After navigating to the destination in the VE, participants completed two memory tests—a pointing task followed by a model building task. The purpose of these tests was to enable us to assess participants’ final representation of the environment, following navigation. Performance on these tests is referred to as memory performance.

**2.4.3.1. Pointing task.** In the pointing task, on each trial, participants were placed directly next to the one of the four landmark buildings of the route and were asked to point to one of the other buildings from that location. The prompt appeared at the top of the screen (e.g., “Point to Harvey House,” while being next to Batty House). To respond, participants were instructed to move a crosshair that appeared in the center of the screen, until it pointed to where they imagined the front door of the building in the prompt to be. Participants could rotate the crosshair in the horizontal plane by moving their mouse and click once to log their response.

The order of locations from which participants pointed matched the order of the buildings along the route, whereas the order of the buildings to which participants pointed was randomized. Once participants had pointed to all three buildings from the first building, they were automatically repositioned at the next building of that route, and pointed to the other three buildings from there. For each trial, performance was assessed by determining the smallest possible angle between the correct answer and the participant’s estimate.

**2.4.3.2. Model building.** For the model building task, participants viewed a blank box on the computer screen with top-down views of each of the four landmark buildings of the route underneath it. Participants were told that the box represented the entire VE they had explored on that route and that they had to place each building where they considered it to be. Participants could drag and drop buildings using their mouse and could adjust their positions as much as necessary. Accuracy on the model-building task was measured using a bidimensional regression analysis (Friedman & Kohler, 2003).

Participants then completed the same procedure (study, navigation, pointing, model building) for a second block involving a different route. After completing this series of tasks for both routes, participants were debriefed. Experimental sessions took about 1.5 hours.

## **2.5. Coding gestures from the study phase**

The gestures in the videos of the study phase were coded in ELAN (Brugman & Russel, 2004). We examined *how much* participants gestured by computing the frequency of their gestures and the proportion of study time they spent gesturing, as well as *how* they gestured by considering the distribution of gesture types and the underlying spatial perspective of these gestures.

### **2.5.1. Gesture frequency**

We determined the number of gestures produced per minute during the study of route directions. Gestures were defined as movements of the hands that depicted semantic content (*representational* gestures) or served discourse or other functions (*nonrepresentational* gestures). Individual gestures

corresponded to distinct *strokes* of gesture—the expressive and dynamic part of the gesture’s execution. For more information about identifying strokes and other phases of gesture execution, as well as for exceptions to this criterion of equating gestures with strokes, please see the Supplemental Methods section in Appendix B.

### 2.5.2. *Gesture duration*

Coding gestures involved identifying not only the gesture’s stroke, but also the onset and offset of the entire gesture’s movement. These time points yielded the duration of each gesture, from which we computed the proportion of study time that was spent gesturing. The Supplemental Methods in Appendix B provide additional details on the segmentation criteria for identifying the onsets and offsets of gestures.

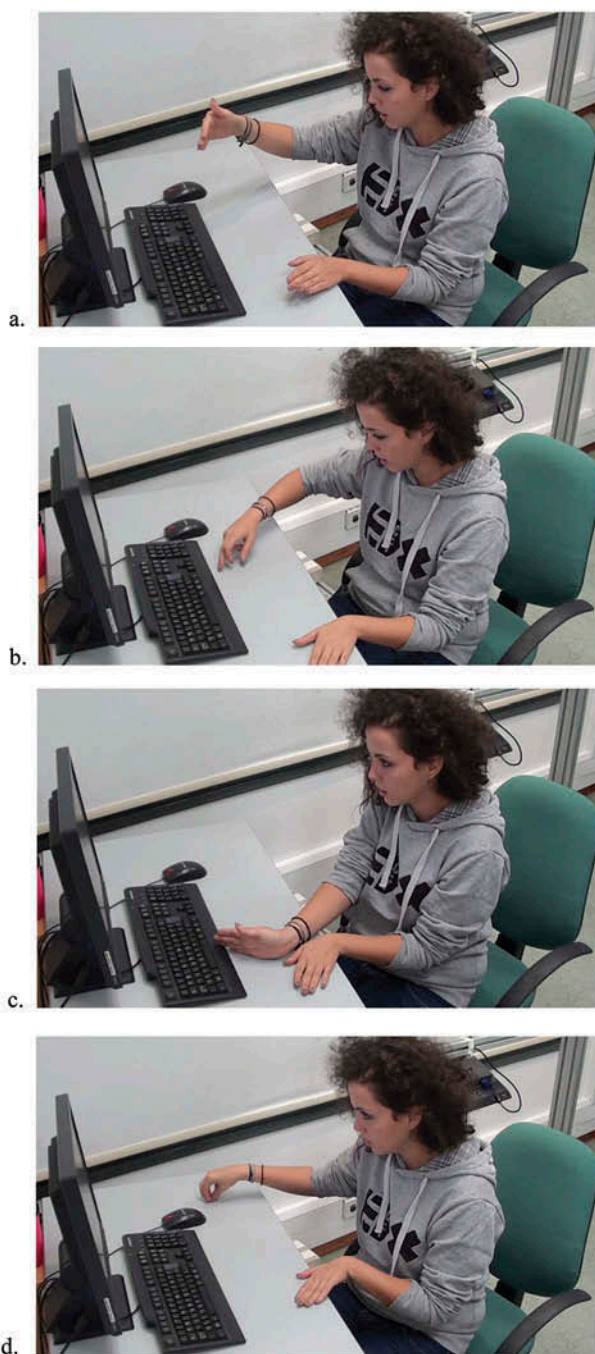
### 2.5.3. *Gesture types and perspective*

Once a gesture was identified and segmented, its type and underlying spatial perspective were identified. This decision involved distinguishing *representational* and *nonrepresentational* gesture types, as the latter did not implicate a spatial perspective. Representational gestures (also known as iconics, McNeill, 1992, or illustrators, Ekman & Friesen, 1969) depict semantic content by virtue of hand-shape, placement, and motion and often represent the movement of characters or properties of objects. Here, representational gestures referred to those gestures encoding spatial content pertinent to the route descriptions. Representational gestures in this study also included abstract pointing gestures, which were used to set up or locate objects in gesture space (see *Location* gestures later).

Because representational gestures here encoded spatial information about the described environment, they were classified as having a route perspective, a survey perspective, or a combination of both perspectives. In a route perspective, the viewer is taken on a mental tour of the environment, with the imagined viewpoint changing within the scene as the viewer’s position changes. In a survey perspective, the imagined viewpoint is stationary and external to the environment (e.g., Emmorey, Tversky, & Taylor, 2000). The spatial perspective of representational gestures was coded strictly based on gesture form, in order to minimize the extent to which coders imputed meaning to hand movements. This approach made it possible to code the spatial perspective of gestures, even when gestures were produced without any accompanying speech, as it was often the case during the study phase.

For *Route* gestures, the directionality of the movement was indicated by one of two required form features: either a palm oriented sideways with fingers typically extended and together (as in Figure 3a) or a pointing handshape moving in 3D space.

For *Survey* gestures, the required form feature was the use of a 2D plane, such as the table or a horizontal plane parallel to the table’s surface (see also



**Figure 3.** Examples of gestures spontaneously produced by Participant 4 while studying route directions, including a gesture from a *Route* perspective encoding a path through movement in 3D space (3a), from a *Survey* perspective encoding a path through tracing on the table (3b), a *Combination* gesture both a *Route* and *Survey* perspective (3c), and a *Location* gesture, encoding the location of a landmark (3d).

Mol, Krahmer, Maes, & Swerts, 2012, for a similar distinction based on form-based features). These gestures represented a path on a 2D plane, for example by tracing a line with the index finger on the table, as seen in Figure 3b).

Moreover, there were representational gestures whose form features captured both spatial perspectives; these were coded as *Combination* gestures. For example, a participant could use the 2D surface of the table to represent the environment, while the orientation, shape, and movement of their gesturing hand reflected the viewer's viewpoint within the environment (e.g., a palm oriented sideways with its edge on the table tracing a path, see Figure 3c).

Finally, *Location* gestures were representational gestures that identified a landmark's location or the navigator's position on a 2D plane (e.g., pointing or tapping on the table with extended finger, clasped fingertips, or the palm, see Figure 3d). Although location gestures can be thought to represent information from an allocentric map-like perspective, we coded them as a separate category, given their distinct form and their indexical semiotic properties. Location gestures can be seen as abstract deictic gestures (Cassell & McNeill, 1991), used to set up or locate characters or objects in gestures space.

Nonrepresentational gestures, which did not encode information pertinent to the routes, included beat gestures—simple “up-and-down” movements of the wrist or fingers that did not encode semantic content—and other discourse gestures (Alibali, Heath, & Myers, 2001; McNeill, 1992), for example when participants enumerated the steps of the route by counting using their fingers.

The dataset also included gestures that were *Ambiguous*; see the Supplemental Methods (Appendix B) for a description of three different levels of ambiguity in gestures (Gesture vs. No Gesture; Representational vs. Nonrepresentational gesture; Route vs. Survey representational gesture) and how these were dealt with.

## 2.6. Coding navigation performance

Using annotating tool ELAN (Brugman & Russel, 2004), the videos from the navigation phase were coded for the following dimensions, referred to as navigation performance.

### 2.6.1. Route duration

This measure captured the time taken to traverse a route. The onset of this duration was operationalized as the first video frame of (typically forward) movement at the origin of the route. Similarly, the offset of this duration was operationalized as the final frame of movement (forward, backward, or lateral) at the destination building of the route.

### 2.6.2. Navigation errors

There were two types of navigation errors.

*Wrong choice point* errors referred to deviations from the route arising from a wrong choice at a decision point, such as a turn, an intersection, a crossroad, or a forked road. Such an error would occur by selecting an incorrect path at a decision point.

*Missed choice points* errors referred to deviations from the route arising from bypassing a landmark. This could happen, for example, when passing by the destination or other landmark buildings (e.g., Tobler Museum in Route 2) without noticing them.<sup>1</sup>

### 2.6.3. Pauses

Pauses were identified as the segments of the video on which the navigator was stationary, without any forward, backward, or lateral movement for two or more frames. The navigator was considered to be stationary, if from frame to frame the video image either remained unchanged or the optic flow suggested a change in heading (i.e., rotation) but not forward, backward, or lateral movement. To control for differences in route duration, we analyzed the proportion of the route's duration that navigators spent pausing (total duration of all pauses / route duration).

## 2.7. Reliability

The first author coded the gestures of all participants, while a second coder coded redundantly the gestures of four participants. In the eight videos coded by both, the coders exhibited high agreement in identifying hand movements that were gestures (the single measure intraclass correlation coefficient (ICC) was .95,  $p < .001$ , using type consistency) and demarcating their duration (ICC = .95,  $p < .001$ ). Agreement was also high for classifying gestures as representational (frequency: ICC = .86,  $p < .01$ ; duration: ICC = .87,  $p < .01$ ), for classifying representational gestures that were from a route perspective (frequency: ICC = .86,  $p < .001$ ; duration: ICC = .98,  $p < .001$ ), a survey perspective (frequency: ICC = .93,  $p < .001$ ; duration: ICC = .79,  $p < .01$ ), a combination of perspectives (frequency: ICC = .94,  $p < .001$ ; duration: ICC = .99,  $p < .001$ ), and for identifying location gestures (frequency: ICC = .89,  $p < .01$ ; duration: ICC = .84,  $p < .001$ ). The first author's coding was used for the analysis.

Reliability for navigation coding was established in Galati et al. (2015), which used the same routes and an identical coding scheme. There was high agreement for determining route duration, deviations from the route, and pause duration between the first author and another coder, who coded

---

<sup>1</sup>Felicitous deviations from the route (e.g., going up to the entrance of Tobler Museum to read its sign) were identified on the basis of the Experimenter's notes and audio recordings of the navigation phase, and were not counted as navigation errors. Backtracking to an earlier location on the path to recover from an error, whether prompted by the experimenter or self-initiated, was also not considered an erroneous deviation.

redundantly 8 videos. That coder proceeded to code the navigation behavior of all videos of the present study.

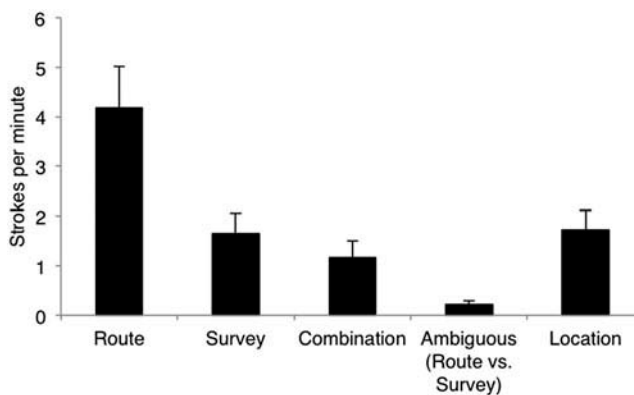
### 3. Results

We first consider the distribution of the gestures that participants spontaneously produced while studying the verbal route directions. Next, we examine the associations between individual ability, navigation, and memory performance. Finally, we examine models that assess the overarching relations among individual ability, gesturing, navigation performance, and memory performance.

#### 3.1. Distribution of gesture types at study

Overall, participants produced 4665 hand movements, 4240 of which were classified unambiguously as gestures. Participants produced 0 to 334 gestures while studying a given route, with a median of 63 gestures. Four individuals (16.67% of participants) produced two or fewer gestures per route (i.e., non-gesturers), with the remaining participants producing 19 gestures or more on average per route.

The majority of the gestures produced (62%) were representational. As shown in Figure 4, route gestures were the most frequent, followed by survey, combination, and location gestures, and lastly by a small number of ambiguous representational gestures (Route vs. Survey). The incidence of these five types of representational gestures differed significantly, as evidenced by a main effect of gesture type on both their frequency,  $F(4, 88) = 8.68, p < .001$ , and their proportional duration,  $F(4, 88) = 5.36, p < .01$ . For both measures,



**Figure 4.** Gesture frequency, in terms of gesture strokes per minute, across the different types of representational gestures: Route, Survey, Combination (both Route and Survey), Ambiguous representational gestures (Route vs. Survey), and Location. Error bars represent standard error of the mean.

route gestures were significantly more frequent and proportionally longer than all other types, ambiguous gestures were less frequent and proportionally shorter than all other types (all  $ps < .05$ ), while survey, combination, and location gestures did not differ significantly from each other (all  $ps > .05$ ).

### 3.2. Individual differences and gesture production

Two main patterns were observed when considering gesture production relative to individual ability. First, participants who reported higher verbal ability (scoring higher on the PVA scale) gestured more frequently (Pearson's  $r = .69, p < .001$ ) and for a greater proportion of the study phase (Pearson's  $r = .72, p < .001$ ). This pattern held for the frequency and duration of route, survey, and location gestures (all  $ps < .05$ ). Second, participants with worse spatial perspective-taking ability (making larger SOT errors), produced ambiguous hand movements that could not be clearly classified as gestures (Ambiguous: Gesture vs. No Gesture) more frequently and for longer durations (for both Pearson's  $r = .45, p < .05$ ). Participants with larger SOT error also produced gestures that were ambiguous in terms of whether they were representational or nonrepresentational for proportionally longer durations of the study phase (Pearson's  $r = .42, p < .05$ ).

### 3.3. Individual differences, navigation and memory performance

Perhaps surprisingly, despite the correlations between some aspects of individual ability and gesturing reported above, individual differences did not correlate with navigation performance, as shown in Table 1. Nevertheless, spatial ability, as captured by SOT error, was significantly correlated with participants' memory performance on the model building task: those with higher spatial perspective-taking ability (smaller SOT) created more accurate model reconstructions (higher  $R^2$ ). Performance on the model building task was also significantly correlated with navigation performance: better navigation performance (indicated by shorter route durations, fewer navigation errors, and less pausing) was associated with significantly more accurate model reconstructions.

Navigation performance was correlated with some aspects of gesturing. For instance, as the duration of route gestures or ambiguous gestures that could not be clearly classified as representational increased (Ambiguous: Representational vs. Nonrepresentational), the proportional duration of pausing during their navigation phase also increased (for the duration of route gestures: Pearson's  $r = .56, p < .01$ ; for the duration of ambiguous gestures: Pearson's  $r = .59, p < .01$ ).



**Table 1.** Bivariate correlations, means, and standard deviations for measures of individual ability, navigation and memory performance for all participants ( $N = 24$ ).

	1	2	3	4	5	6	7	8	9	$M$	$SD$
1. SOT (Error)	–									47.4	30.9
2. PSAS	-.42*	–								5.0	.7
3. PVA	.14	.21	–							4.5	.7
4. SBSOD–CY	-.48*	.68**	-.06	–						4.8	1.0
5. Route duration	.26	-.01	.03	-.01	–					298.9	122.6
6. Navigation errors	.34	-.09	-.01	-.14	.85**	–				1.9	1.4
7. Proportion pausing	.30	.06	.10	.04	.86**	.61**	–			.3	.2
8. Pointing (Error)	.34	-.11	-.19	-.31	-.03	.22	-.01	–		42.7	36.7
9. Model building ( $R^2$ )	-.53**	-.07	-.38	.09	-.42*	-.49*	-.49*	-.23	–	.8	.1

For measures labeled (Error) higher values indicate worse performance, whereas for the other measures higher values indicate better performance. SOT = Spatial Orientation Test; PSAS = Philadelphia Spatial Ability Scale; PVA = Philadelphia Verbal Ability Scale; SBSOD = Santa Barbara Sense of Direction. Route duration is the mean duration for navigating each route (in secs), navigation errors is the mean total number of wrong and missed choice point errors per route, and model building ( $R^2$ ) is the mean of  $R^2$  for the two routes.

\*  $p < .05$ , \*\*  $p < .01$ .

### 3.4. Modeling the relations among gesturing, navigation, and memory

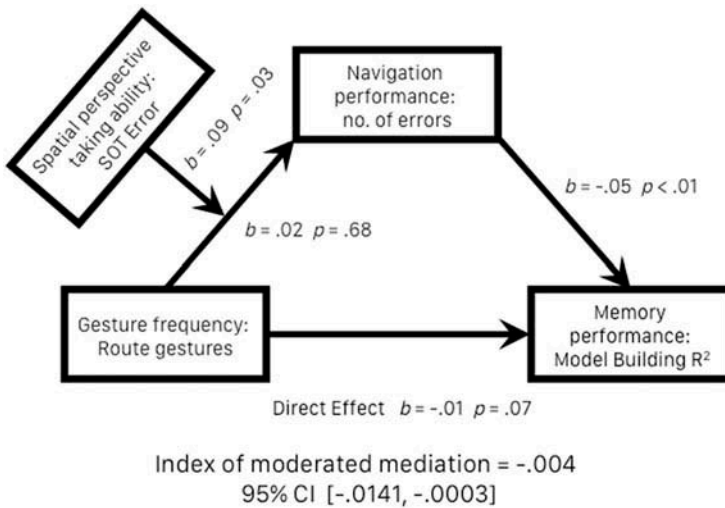
Given the challenges in interpreting the plethora of correlations between the metrics of individual ability, gesturing, navigation performance, and memory performance, we used a single analytical model to assess an overarching, theoretically motivated relationship between behavior across the different phases of the study.

Specifically, we reasoned that the spatial representation of the environment that participants used in the navigation and testing phases of the experiment was not necessarily the same. During navigation, participants were presumably guided by an *initial* representation that they had constructed at study with linguistic descriptions as input and with gestures potentially elaborating that representation. On the other hand, during the memory tests, as suggested by Figure 2, participants accessed their *final* representation of the environment, which had been likely enriched and updated by visual information from the virtual environment experienced during navigation.

In light of this, we conceptualized navigation performance as mediating the relationship between gesturing at study and memory performance. Moreover, given the extant literature we have reviewed, we hypothesized a potential interaction between individual ability and gesturing, with the effect of gesturing on performance potentially differing depending on individual ability. In terms of our model, we conceptualized individual ability as moderating the influence of gesturing on navigation performance. Thus, putting these pieces together, we examined a model according to which individual ability moderates the relationship between gesturing and navigation performance, and navigation performance in turn mediates the relationship

between gesturing and memory performance. This conceptual model is illustrated in Figure 5.

We used Hayes's (2012) PROCESS computational tool within the SPSS environment to conduct conditional process analyses, which incorporate moderation and mediation in a single integrated analytical model. We analyzed separately findings with gestures from a route vs. survey perspective, as those could reflect different underlying strategies used at study<sup>2</sup>. We focused on models with SOT error as the moderator given our theoretical interest in the potential interaction between spatial perspective-taking ability and the spatial perspective of gestures. Our models were also constrained to those with the correlation coefficient squared ( $R^2$ ) of the model building task as the outcome measure, because correlational analyses seen in Table 1 (as well as those reported in Galati et al., 2015) suggested that SOT error and  $R^2$  would be good indices of spatial ability and memory performance, respectively. Moreover, in the present study, the model building  $R^2$  (but not pointing error) was significantly correlated with all three measures of navigation performance (all  $ps < .05$ ), and was significantly correlated only with SOT error ( $p < .01$ ) among the individual ability measures.



**Figure 5.** Model of the conditional effect of route gesture frequency on memory performance (the correlation coefficient squared ( $R^2$ ) on the model building task), with the navigation performance (the number of deviation errors during navigation) as the mediator of that relationship, and individual spatial perspective-taking ability (the standardized SOT error) as the moderator of the relationship between gesturing and navigation.

<sup>2</sup>In models with the frequency or duration of representational gestures (combining the categories of route, survey, location, and ambiguous: route vs. survey gestures), there was no evidence of moderated mediation, whether in terms of the index of moderated mediation or in terms of the conditional indirect effect of gesturing.

We only report those models that yielded fairly definitive evidence of moderated mediation. Such evidence came from two sources: first, from the *index of moderated mediation* (Hayes, 2015), and second, from examining the *conditional indirect effect* of the predictor (Hayes, 2012). The index of moderated mediation quantifies the association between an indirect effect and a moderator, followed by an inference as to whether this parameter is different from zero based on a bootstrap confidence interval of its estimate (Hayes, 2015).

When this estimate is different from zero, this suggests that the indirect effect is moderated: that conditional indirect effects of the predictor (i.e., a measure of gesturing) at some different values of the moderator (i.e., spatial ability) will be significantly different from one another. In order to probe further into this moderated mediation, we examine the conditional indirect effect by obtaining estimates of the indirect effect of the predictor conditioned at different values of the moderator. Changes in the size or direction of indirect effect of the predictor at different levels of the moderator provide additional evidence of moderated mediation and clarify its nature. We explain this in more detail below when presenting the results of Table 2.

### 3.4.1. Models with route gestures as the predictor

We begin with findings from the model in Figure 5, which we will describe in some detail. As shown, the model had the frequency of route gestures produced at study as the predictor, the  $R^2$  metric of the model building task as the outcome measure, the number of errors during navigation as the mediator of the relationship between gesturing and memory performance, and SOT error as the moderator of the relationship between gesturing and navigation.

Results showed that there was significant evidence for moderated mediation, as indicated by the index of moderated mediation, whose confidence interval did not contain zero (parameter estimate =  $-.004$ ,  $SE = .003$ , 95% CI

**Table 2.** The conditional indirect effect of gesturing (frequency of Route gestures) on memory performance ( $R^2$  of model building task) through navigation performance (number of deviation errors) for various levels of spatial ability (ZSOT error).

ZSOT error	Effect ( $\omega$ )	$se_{\omega}$	95% Bootstrap CI
-1.1558	.0040	.0041	[-.0031, .0139]
-.7702	.0023	.0034	[-.0039, .0094]
-.4497	.0009	.0030	[-.0051, .0071]
.7327	-.0042	.0041	[-.0195, .0005]
1.2361	-.0065	.0053	[-.0289, -.0005]

The values for ZSOT error are the 10th, 25th, 50th, 75th, and 90th percentiles. The  $\omega$  is the point estimate of the conditional indirect effect for a given value of the moderator (ZSOT error), along with its bootstrap standard error and confidence intervals.

[-.0141, -.0003]). This suggested that route gestures had an indirect effect on memory performance through navigation that depended on spatial perspective-taking ability.

To understand further this moderated mediation, we examined the *conditional indirect effect* of the predictor on the outcome variable. Note that if a predictor's (direct or indirect) effect on the outcome measure is moderated, this effect cannot be quantified with a single numerical estimate, because the effect differs in size or strength as a function of the moderator variable (Hayes, 2013). Instead, the discussion of the predictor's effect on the outcome measure (i.e., of gesturing on memory performance) must be conditioned on the moderator (i.e., individual spatial ability). To do so, we consider estimates of the conditional indirect effect of the predictor at different levels of the moderator.

Table 2 presents the point estimate of conditional indirect effect ( $\omega$ ) of the frequency of route gestures on memory performance ( $R^2$  of model building task) through navigation performance (number of deviation errors) at five different percentile points of spatial perspective-taking ability (ZSOT error). As shown, the only confidence interval of the estimate of this conditional indirect effect ( $\omega$ ) that excluded zero was the one for the group representing the top 10% of mean SOT error, in the last row of the table. That is, there was a significant negative relationship between route gesturing and memory performance (given the negative sign of the estimate) only for those with the lowest spatial ability.

In terms of the rest of the model, there was evidence that SOT error influenced the relationship between route gestures and navigation, accounting for 24% of the variance in navigation errors, as reflected by a significant interaction term between SOT error and the frequency of route gestures was significant ( $b = .088$ ,  $SE\ b = .037$ ,  $t(3) = 2.37$ ,  $p < .05$ ,  $R^2 = 24\%$ )<sup>3</sup>. Finally, navigation performance significantly predicted memory, as shown in Figure 5 ( $b = -.0496$  95% CI [-.0806, -.0185],  $t(2) = 3.32$ ,  $p < .01$ ,  $R^2 = 37\%$ ): as navigators made more errors during the navigation phase, they produced more distorted reconstructions of the environment (i.e., lower correlation coefficient squared  $R^2$  on the model building task). The direct effect of the frequency of route gestures on the  $R^2$  of the model building task was negative and nonsignificant ( $p = .07$ ).

In sum, findings from this model suggest that indirect effect of route gestures on memory performance changed across different levels of SOT. For low spatial ability individuals, increased route gesturing was associated with worse memory performance (through navigation). For the remaining groups, this indirect effect was not significant.

<sup>3</sup>Note that using standardized variables (namely, ZSOT) complicates the interpretation of the value of regression coefficients, because standardization results in estimates of the variable's effect that have to be interpreted in terms of differences between cases that differ by one standard deviation (vs. one unit in the original metric of measurement). These regression coefficients should be interpreted only in terms of their sign (Hayes, 2013).

### 3.4.2. Models with survey gestures as the predictor

For models with measures of survey gesturing as the predictor in the conceptual model of Figure 5, there was fairly definitive evidence<sup>4</sup> of the hypothesized moderated mediation, even in the absence of significant moderation of spatial perspective-taking ability on the effect of gesturing on navigation (i.e., when the interaction term capturing the moderation of SOT did not reach significance). Moderated mediation was indicated through changes in the conditional indirect effect of survey gesturing on memory performance (through navigation performance) across different levels of spatial perspective-taking ability.

There were two converging models demonstrating that the mediated effect of survey gesturing on memory performance changed across different levels of spatial ability. In one model, there was a significant positive relationship between the frequency of survey gestures and the  $R^2$  of the model building task through navigation (through the duration of navigating the route), but only for those with the lowest spatial perspective-taking ability (at the top 25% of mean SOT error). For these low spatial ability individuals, producing more survey gestures predicted (through navigation performance) more accurate reconstructions of the environment (at the top 75<sup>th</sup> percentile of SOT error: effect = .017,  $SE$  = .009, 95% Bootstrap CI [.002, .043], and at the top 90<sup>th</sup> percentile: effect = .020,  $SE$  = .011, 95% Bootstrap CI [.001, .048]). In a similar model with the number of navigation errors as the mediator, the same pattern was observed: low spatial ability individuals produced more accurate reconstructions when having produced more survey gestures (at the top 75<sup>th</sup> percentile of SOT error: effect = .029,  $SE$  = .011, 95% Bootstrap CI [.004, .061], and at the top 90<sup>th</sup> percentile: effect = .023,  $SE$  = .014, 95% Bootstrap CI [.002, .063]).

## 4. Discussion

We set out to examine whether, during route learning from verbal descriptions, the navigators' individual ability interacts with their co-thought gesturing to predict the accuracy of their memory representation for the navigated environment. The findings suggest that gesturing while studying route directions does not confer a global advantage for subsequent memory performance. Instead, gesturing predicted (through the navigation experience) the accuracy of the memory representations of a specific group of

<sup>4</sup>A nonsignificant interaction (i.e., a confidence interval for the regression of the interaction term of the moderation that includes zero) does not imply that the indirect effect is not moderated by individual ability, because that interaction does not quantify the relationship between the moderator and the indirect effect. A bootstrap confidence interval for the index of moderated mediation that does not include zero provides more direct and definitive evidence of moderation of the indirect effect of the predictor on the outcome than a test of moderation of one of its paths (Hayes, 2015).

individuals—those with low spatial perspective-taking ability. What's more, the relationship between gesturing and memory performance for these low spatial ability individuals differed according to the perspective of their gestures. For low spatial ability individuals (but not for other groups), increased gesturing from a route perspective was associated with *less* accurate memory representations, whereas increased gesturing from a survey perspective was associated with *more* accurate memory representations mediated by the navigation experience.

These findings extend indirect evidence from previous studies that have suggested that individual ability interacts with gesturing in some tasks. For example, there is evidence of a nonlinear relationship between individual ability and gesturing, whereby those with high and low phonemic fluency produce more representational gestures than those with average phonemic fluency (Hostetter & Alibali, 2007). The possibility of an interaction between individual ability and gesturing is also supported by findings that individuals with low working-memory capacity benefit from gesturing in a dual-task, whereas high working-memory individuals do not (Marstaller & Burianova, 2013). To our knowledge, our study is the first to demonstrate an interaction between individual ability and gesturing on performance in a spatial task, and additionally to demonstrate that the relationship between gesturing and performance differs according to the type of gestures produced.

An observation relevant to the profile of low spatial ability individuals, who were driving the observed patterns here, is that they gestured less clearly: as spatial ability decreased, individuals were more likely to gesture ambiguously, both in terms of whether their hand movements could be classified as gestures (vs. not gestures) and whether their gestures were representational (vs. nonrepresentational). Another approach toward validating this finding about gesture clarity would involve coders (or another set of participants) providing perceptual judgments about the precision of gestures (e.g., see Galati & Brennan, 2014).

Here, the increase in gestures that were individually identified and classified as ambiguous may suggest that individuals with lower spatial perspective-taking ability constructed “fuzzier,” less accurate spatial representations. It's not possible to disentangle whether such ambiguous gestures reflected the less accurate representations of those gesturers or contributed to shaping these representations to be less accurate—both processes may have been at play. What we can state is that low spatial ability individuals were more inclined to gesture ambiguously, and when they *did* produce clear representational gestures these didn't predict their memory performance uniformly.

Importantly, the spatial perspective of clear, representational gestures mattered for those low spatial ability individuals: survey gestures were associated with improved performance, while route gestures were associated with a detriment in performance. These distinct patterns for route and survey

gestures support the possibilities we had laid out about their respective contribution to spatial learning. We had hypothesized that survey gestures may be more useful to spatial learning than route gestures because, in addition to representing directional turns, they allow the externalization and appraisal of global relationships between directional turns and other features of the environment (including landmarks). Our findings broadly support this possibility, at least for those with poorer spatial perspective-taking ability. Nevertheless, given the correlational nature of the study, we do not have definitive evidence for a causal effect of gestures.

Beyond semiotic differences between route and survey gestures, it's also possible that any differential contribution of these gestures to spatial learning stems from the complementarity between the perspective of the gestures and the perspective of the linguistic descriptions provided. Here, the linguistic descriptions involved only a route perspective, similar to several other contexts and interfaces in which directions are provided from the navigator's perspective (e.g., GoogleMaps directions). Spatial information in these route directions was expressed as numbered steps that could be reduced to a sequence of left turns, right turns, and straight segments. These serialized bits of information did not describe directly the global spatial structure of the environment.

Global relationships could only be inferred from the route descriptions, which can be demanding in terms of cognitive resources (e.g., Pazzaglia, De Beni, & Meneghetti, 2007; Brunyé & Taylor, 2008). Similarly, gestures from a route perspective also expressed sequential bits of directional information, echoing the information in the text without adding any information over and above the text. In contrast, gestures from a survey perspective, encoded additional information by representing global relationships of the environment inferred from route descriptions (e.g., interrelationships between the route and landmarks). This additional information about global relationships, inferred from route directions and externalized through survey gestures, may have made a difference at test for low spatial ability individuals.

Previous work has also underscored that the complementarity of perspectives across representational formats can benefit spatial performance (Brunyé, Rapp, & Taylor, 2008). Brunyé and colleagues showed that, when studying route-based description, having an accompanying representation from a survey perspective (namely, viewing a map) contributed to more accurate inferences about spatial relationships. The researchers proposed that this benefit in performance arises because, by having to integrate different spatial perspectives, people construct a more flexible representation of the environment.

The idea that representational flexibility results from the integration of spatial perspectives is also related to the view that, as people accumulate spatial knowledge about a novel environment, they recruit concurrently—

and from the earliest exposure to the environment—route knowledge and metric configural knowledge (Montello, Waller, Hegarty, & Richardson, 2004). In the present study, although participants were not provided with a survey representation, they could construct one through survey gestures. The process of linking directional information from route directions to their externalized gesture model may have contributed to a more flexible representation that was especially useful to poor spatial perspective-takers. In contexts of spatial learning that permit gesturing, survey gestures can be thought to serve as a bridge to configural knowledge about the environment during its initial encoding, even when the input is purely linguistic and does not contain metric information.

It is still an open question what the underlying reasons are for the potentially differential contribution of route and survey gestures to spatial performance here. The intriguing possibilities that our work hints at are: that the patterns observed are driven by the representational flexibility arising from the complementarity of perspectives broadly, that they are driven specifically by inferences about global relationships when survey gestures complement route descriptions, or that they are driven by the semiotic properties of survey gestures alone. These possibilities require systematic probing in future empirical work.

Given the distinct predictions they afford, they can be assessed through studies that manipulate the perspective of the linguistic material (i.e., having directions from a route vs. survey perspective), and perhaps also instructions to gesture from a route vs. survey perspective. These manipulations can clarify whether the potential benefit of survey gestures we have observed is best explained by the complementarity of perspectives broadly (in which case route gestures paired with survey descriptions would confer a benefit comparable to survey gestures paired with route descriptions, relative to pairings of the same perspective), by the complementarity of the specific pairing of survey gestures with route descriptions, or by the perspective of survey gestures alone (independently of the perspective of the linguistic descriptions).

Another caveat here is that having linguistic directions as the sole input may have also contributed to the significant positive correlation observed between participants' self-reported verbal ability and the production of representational gestures. Verbal ability (and perhaps, by extension, gesturing) may have mattered less if directions included imagistic representations. Interestingly, the opposite relationship between verbal ability and gesturing has been reported elsewhere, with individuals with lower verbal working memory gesturing more frequently (Gillespie et al., 2014). Beyond differences in the verbal ability measures used (scores from a self-reported scale of verbal ability vs. from standardized tasks assessing verbal working memory), one important methodological difference that may account for this discrepancy is that participants here had to process linguistic input (whereas in Gillespie



et al., 2014, the elicitation stimuli were nonverbal cartoons) and did not have to produce speech (whereas in Gillespie et al., 2014, participants produced linguistic descriptions). Co-thought gestures produced during language comprehension and (co-speech) gestures produced during language production may rely differently on verbal ability.

Because there was no correlation between verbal and spatial ability in this sample, studies with larger sample sizes can permit distinguishing groups that differ systematically in these abilities to better assess how their spatial learning outcomes are influenced by gesturing. Our study's relatively small sample size does not permit examining how gesturing patterns across different breakdowns of spatial and verbal ability (e.g., high-high, high-low, low-high, low-low). Studies with larger sample sizes can also permit examining differences in preference in gestural perspective (i.e., those who prefer gesturing from a survey relative to a route perspective, and vice-versa) against their performance. Some exploratory analyses we conducted on the present data did not afford any additional insights. For instance, we found that as preference for a route over a survey perspective increased in gesture, individuals paused more and took longer to complete the routes during navigation, in line with our reported findings.

Given our methodological decision to not manipulate gesturing, but rather to allow participants to gesture spontaneously, we have been careful to not frame the relationship between gesturing and spatial performance (on navigation and spatial memory) as a causal one. Instructed vs. spontaneous gesturing each have their merits as methodological decisions, with the former affording conclusions about causality and the latter permitting the more naturalistic observation of gestural behavior. In Galati et al. (2015), where we did in fact manipulate gesturing through instructions, by asking participants to gesture or not gesture while studying route descriptions, we did not find an overall benefit of forced gesturing (vs. gesture restriction) on memory or navigation performance.

However, because in that work we did not examine the frequency and distribution of the gestures produced, it remains unclear whether spontaneous gesturing and instructed gesturing from a particular perspective confers the same benefits for spatial learning. Other work examining the effect of co-speech gestures in nonspatial tasks has shown no reliable differences between forced and spontaneous gesturing (e.g., Goldin-Meadow, Nusbaum, Kelly, & Wagner, 2001). In either case, the benefit of interventions using gesture warrants further investigation in the context of spatial learning.

Some initial gesture-based interventions in the spatial domain have been shown to be promising. For example, promoting the use of gestures when students reason and communicate about 3D spatial relationships has resulted in improved performance in measures of “penetrative thinking”—the ability to visualize the interior of 3D structures (Atit, Gagnier, & Shipley, 2015). Future work may examine whether promoting the use of survey gestures

when encoding spatial information benefits performance, particularly for those with low spatial ability.

The findings add to the sparse literature on co-thought gestures, underscoring that people spontaneously recruit these gestures during route learning. When learning route directions, gesturing appears to have a selective effect that depends on the gesturers' spatial perspective-taking ability and on the perspective of the gestures they produce. Gesturing from a survey perspective, whether due to its complementarity with the perspective of route descriptions or due to linking the constructed spatial representation to the immediate gesture space, is associated with improved performance for navigators with low spatial ability.

## Acknowledgments


We are grateful to Victor Schinazi for designing the virtual environment used in the studies, to Michalis Michaelides for useful discussions on the data analysis, to Philippos Panagiotou for assistance with data collection, and to Georgia Leonidou, Elena Stylianopoulou and Artemis Stefani for assistance with coding. We also thank Dan Montello and three anonymous reviewers for suggestions that significantly improved this article.

## Funding

This material was supported by a SAVI grant from NSF to A.G., through the Spatial Intelligence and Learning Center (SILC) SBE 1041707, and by ERC-2007-StG 206912-OSSMA from the European Research Council to M.A.

## ORCID

Alexia Galati  <http://orcid.org/0000-0001-7593-6629>

Nora S. Newcombe  <http://orcid.org/0000-0002-7044-6046>

## References

- Alibali, M. W. (2005). Gesture in spatial cognition: Expressing, communicating, and thinking about spatial information. *Spatial Cognition & Computation*, 5, 307–331.
- Alibali, M. W., Heath, D. C., & Myers, H. J. (2001). Effects of visibility between speaker and listener on gesture production: Some gestures are meant to be seen. *Journal of Memory and Language*, 44, 169–188.
- Alibali, M. W., Spencer, R. C., Knox, L., & Kita, S. (2011). Spontaneous gestures influence strategy choices in problem solving. *Psychological Science*, 22, 1138–1144.
- Allen, G. (2003). Gestures accompanying verbal route directions: Do they point to a new avenue for examining spatial representations? *Spatial Cognition and Computation*, 3, 259–268.
- Atit, K., Gagnier, K., & Shipley, T. F. (2015). Student gestures aid penetrative thinking. *Journal of Geoscience Education*, 64, 66–72.

- Avraamides, M. N., & Kelly, J. W. (2008). Multiple systems of spatial memory and action. *Cognitive Processing*, 9, 93–106.
- Brugman, H., & Russel, A. (2004). Annotating Multi-media / Multi-modal resources with ELAN. In M. Lino, M. Xavier, F. Ferreira, R. Costa, & R. Silva (Eds.), *Proceedings of the 4th International Conference on Language Resources and Language Evaluation (LREC 2004)* (pp. 2065–2068). Paris: European Language Resources Association.
- Brunyé, T. T., Rapp, D. N., & Taylor, H. A. (2008). Representational flexibility and specificity following spatial descriptions of real-world environments. *Cognition*, 108, 418–443.
- Brunyé, T. T., & Taylor, H. A. (2008). Working memory mechanisms in developing spatial mental models from survey and route descriptions. *Journal of Memory and Language*, 58, 701–729.
- Cassell, J., & McNeill, D. (1991). Gesture and the poetics of prose. *Poetics Today*, 12, 375–404.
- Chu, M., & Kita, S. (2011). The nature of gestures' beneficial role in spatial problem solving. *Journal of Experimental Psychology: General*, 140, 102–115.
- Chu, M., Meyer, A., Foulkes, L., & Kita, S. (2014). Individual differences in frequency and salience of speech-accompanying gestures: the role of cognitive abilities and empathy. *Journal of Experimental Psychology: General*, 143, 694–709.
- de Ruiter, J. P. A. (2000). The production of gesture and speech. In D. McNeill (Ed.), *Language and gesture: Window into thought and action* (pp. 284–311). Cambridge, UK: Cambridge University Press.
- Ekman, P. & Friesen, W. V. (1969). The repertoire of nonverbal behavior: Categories, origins, usage, and coding. *Semiotica*, 1, 49–80.
- Emmorey, K., Tversky, B., & Taylor, H. (2000). Using space to describe space: Perspective in speech, sign, and gesture. *Spatial Cognition and Computation*, 2, 157–180.
- Friedman, A., & Kohler, B. (2003). Bidimensional regression: Assessing the configural similarity and accuracy of cognitive maps and other two-dimensional data sets. *Psychological Methods*, 8, 468–491.
- Galati, A., & Brennan, S. E. (2014). Speakers adapt gestures to addressees' knowledge: Implications for models of co-speech gesture. *Language, Cognition and Neuroscience*, 29, 435–451.
- Galati, A., Weisberg, S., Newcombe, N., & Avraamides, M. N. (2015). Individual differences in spatial ability influence the effect of gesturing on navigation and spatial memory. In G. Ferré & M. Tutton (Eds.), *Proceedings of gesture and speech in interaction – 4<sup>th</sup> edition (GESPIN 4)* (pp. 119–124). Nantes, France.
- Gillespie, M., James, A. N., Federmeier, K. D., & Watson, D. G. (2014). Verbal working memory predicts co-speech gesture: Evidence from individual differences. *Cognition*, 132, 174–180.
- Goldin-Meadow, S. (2011) Learning through gesture, *Wiley Interdisciplinary Reviews (WIREs): Cognitive Science*, 2, 595–607.
- Goldin-Meadow, S., Nusbaum, H., Kelly, S., & Wagner, S. (2001). Explaining math: Gesturing lightens the load. *Psychological Science*, 12, 516–522.
- Graham, J. A., & Argyle, M. (1975). A cross-cultural study of the communication of extra-verbal meaning by gestures. *International Journal of Psychology*, 10, 57–67.
- Hayes, A. F. (2012). PROCESS: A versatile computational tool for observed variable mediation, moderation, and conditional process modeling [White paper]. Retrieved from <http://www.afhayes.com/public/process2012.pdf>
- Hayes, A. F. (2013). *Introduction to mediation, moderation, and conditional process analysis*. New York, NY: The Guilford Press.
- Hayes, A. F. (2015). An index and test of linear moderated mediation. *Multivariate Behavioral Research*, 50, 1–22.

- Hegarty, M., Crookes, R. D., Dara-Abrams, D., & Shipley, T. F. (2010). Do all science disciplines rely on spatial abilities? Preliminary evidence from self-report questionnaires. In C. Hölscher, T. F. Shipley, M. O. Belardinelli, J. A. Bateman, & N. S. Newcombe (Eds.), *Lecture Notes in Computer Science: Spatial Cognition VII: International Conference, Spatial Cognition 2010, Mt. Hood/Portland, OR, USA, August 15-19, 2010: Proceedings* (Vol. 6222, pp. 85–94). Berlin, Germany: Springer.
- Hegarty, M., Richardson, A. E., Montello, D. R., Lovelace, K., & Subbiah, I. (2002). Development of a self-report measure of environmental spatial ability. *Intelligence*, *30*, 425–447.
- Hegarty, M., & Waller, D. (2004). A dissociation between mental rotation and perspective-taking spatial abilities. *Intelligence*, *32*, 175–191.
- Hostetter, A. B., & Alibali, M. W. (2007). Raise your hand if you're spatial: Relations between verbal and spatial skills and representational gesture production. *Gesture*, *7*, 73–95.
- Hostetter, A. B., & Alibali, M. W. (2008). Visible embodiment: Gestures as simulated action. *Psychonomic Bulletin and Review*, *15*, 495–514.
- Hostetter, A. B., & Alibali, M. W. (2010). Language, gesture, action! A test of the gesture as simulated action framework. *Journal of Memory and Language*, *63*, 24–257.
- Jamalian, A., Giardino, V., & Tversky, T. (2013). Gestures for thinking. In M. Knauff, M. Pauen, N. Sebanz, & I. Wachsmuth (Eds.), *Proceedings of the 35<sup>th</sup> Annual Conference of the Cognitive Science Society* (pp. 645–650). Austin, TX: Cognitive Science Society.
- Kita, S., & Özyürek, A. (2003). What does cross-linguistic variation in semantic coordination of speech and gesture reveal? Evidence for an interface representation of spatial thinking and speaking. *Journal of Memory and Language*, *48*, 16–32.
- Krauss, R. M., Chen, Y., & Gottesman, R. (2000). Lexical gestures and lexical access: A process model. In D. McNeill (Ed.), *Language and gesture* (pp. 261–283). Cambridge, UK: Cambridge University Press.
- Marstaller, L., & Burianová, H. (2013). Individual differences in the gesture effect on working memory. *Psychonomic Bulletin & Review*, *20*, 496–500.
- McNeill, D. (1992). *Hand and mind: What gestures reveal about thought*. Chicago, IL: University of Chicago Press.
- Melinger, A., & Kita, S. (2007). Conceptualisation load triggers gesture production. *Language and Cognitive Processes*, *22*, 473–500.
- Mol, L., Krahmer, E., Maes, A., & Swerts, M. (2012). Adaptation in gesture: Converging hands or converging minds? *Journal of Memory and Language*, *66*, 249–264.
- Montello, D. R., Waller, D., Hegarty, M., & Richardson, A. E. (2004). Spatial memory of real environments, virtual environments, and maps. In G. L. Allen (Ed.), *Human spatial memory: Remembering where* (pp. 251–285). Mahwah, NJ: Lawrence Erlbaum Associates.
- Pazzaglia, F., De Beni, R., & Meneghetti, C. (2007). The effects of verbal and spatial interference in the encoding and retrieval of spatial and non-spatial texts. *Psychological Research*, *71*, 484–494.
- Pazzaglia, F., & Taylor, H. A. (2007). Perspective, instruction, and cognitive style in spatial representation of a virtual environment. *Spatial Cognition and Computation*, *7*, 349–364.
- Sassenberg, U., Foth, M., Wartenburger, I., & van der Meer, E. (2011). Show your hands—Are you really clever? Reasoning, gesture production, and intelligence. *Linguistics*, *49*, 105–134.
- Schinazi, V. R., Nardi, D., Newcombe, N. S., Shipley, T. F., & Epstein, R. A. (2013). Hippocampal size predicts rapid learning of a cognitive map in humans. *Hippocampus*, *23*, 515–528.
- Shimi, A., & Avraamides, M. (2008). The construct validity of the Santa Barbara Sense of direction scale in the Greek population. *International Journal of Psychology*, *43*, 119.

- So, W. C., Ching, T. H.-W., Lim, P. E., Cheng, X., & Ip, K. Y. (2014). Producing gestures facilitates route learning. *PLoS ONE*, 9(11), e112543. doi:10.1371/journal.pone.0112543
- Weisberg, S. M., Schinazi, V. R., Newcombe, N. S., Shipley, T. F., & Epstein, R. (2014). Variations in cognitive maps: Understanding individual differences in navigation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 40, 669–682.
- Wesp, R., Hesse, J., Keutmann, D., & Wheaton, K. (2001). Gestures maintain spatial imagery. *American Journal of Psychology*, 114, 591–600.

## Appendix A

### A.1. Description of Route 1

- (1) You begin with having **Batty House** on your right.
- (2) Continue straight and make a left on the main road.
- (3) Follow this road, moving straight ahead.
- (4) Take the next turn to the left.
- (5) Go straight, passing **Harvey House** on your right.
- (6) At the end of this road, you will see a tree in the middle of the street. From there, make a right turn.
- (7) Follow this road, and take the first turning to the left.
- (8) Go straight, and at the end of the road make a right turn.
- (9) Immediately, you will see a fork in the road in front of you. At the forked road, go left.
- (10) Continue straight, until you come to an intersection.
- (11) At the intersection, you will see **Snow Church** in front of you. From there, make a left turn.
- (12) Continue straight until you reach **Golledge Hall** on your left.

## Appendix B

### B.1. Supplemental methods for gesture coding

#### B.1.1. Identifying gestures

As pointed out in the Method, gesture frequency was computed on the basis of gesture strokes. Strokes were considered to be the expressive and motorically dynamic part of the gesture's execution. For representational gestures, strokes are thought to bear the gesture's semantic content (McNeill, 1992). In terms of their execution, strokes are optionally preceded by a preparation phase, during which the hands move from rest towards the space where the stroke is executed, and a retraction phase, during which the hands return to rest.

In our coding, there were two exceptions to identifying gestures on the basis of a dynamic stroke phase; these involved gestures with a pointing handshape and tracing gestures. For gestures involving a pointing handshape, instead of identifying a stroke, we identified when the extended finger reached its apex, achieving maximum deviation from the movement's onset. For tracing gestures, where extended fingers traced a line (e.g., on the table's surface or a 2D plane in gesture space) with a relatively constant speed, we segmented them by identifying changes in the dynamics of the movement (e.g., deceleration and acceleration, or changes in the direction of the movement).

### B.1.2. *Segmenting and classifying gestures*

The segmentation criteria for identifying a gesture's onset and offset were the following. If the hands were at rest, the onset of a gesture was defined as the first frame of the video on which the hands lifted from rest (i.e., the first frame of the preparation phase). If the hands returned to rest, the offset of the gesture was defined as the first frame on which movement ceased (i.e., the end of the retraction phase). If gestures were interlinked without the hands returning to rest, the transition between one gesture to the next was identified on the basis of the dynamics of the hand movement (e.g., a change in direction). For two-handed gestures that involved asynchronous movement of the hands, the onset was determined by the movement of the hand moving first and the offset by the final movement of the hand returning to rest last.

Preceding or following the stroke phase, there could also be a gesture *hold*, whereby one or both hands remained stationary. For instance, participants could maintain active the hand-shape of a previous gesture, by "holding" active the location of a landmark with fingertips (or palm) on the table. Such post-stroke holds could be maintained by one hand, while the other hand continued to gesture. Post-stroke holds were excluded from the total gesturing duration, although they were still annotated in a separate tier in the ELAN interface.

Importantly, when coding the perspective of gestures, in the cases where one hand remained in hold while the other continued to gesture, the perspective of the stationary hand (e.g., Location for a left palm on the table maintaining the location of a landmark) was not factored into the judgment made on the perspective of the gesture produced by the other hand (i.e., if the right hand produced a Route gesture on a 3D plane off the table, it was classified as Route gesture).

### B.1.3. *Coding ambiguous gestures*

Because gestures were coded on the basis of their form and often in the absence of speech, there were instances in which they could not be clearly classified as belonging to particular category and they were therefore ambiguous.

There were three levels of ambiguity: (1) whether a hand movement was a gesture or not, (2) whether a gesture was representational or nonrepresentational (e.g., whether a finger tap was a representational gesture from a survey perspective locating a landmark, or a nonrepresentational beat gesture), and (3) whether a representational gesture was from a route or a survey perspective.

About 9% (425 out of 4665) of participants' hand movements fell into the first category (of not being clearly classified as gestures or not) and were excluded from subsequent computations of the measures of overall gesture frequency.