1

Thinking with the body: conceptual integration through gesture in multiviewpoint model construction

1 Introduction

Educational researchers in the fields of science, technology, engineering, and math (STEM) have recently come to recognize the value of teaching students to construct, critique, revise, and apply models (Lehrer and Schauble 2006; Schwarz et al. 2009). A model represents an object, a phenomenon, or an idea (Gilbert, 2000) through analogy: a mapping between the target to be modeled (e.g. the solar system) and the source materials that create the model (e.g. wire, styrofoam, tape). In school, national standards for math and science emphasize the benefits of creating and reasoning from models (National Research Council 2002; National Council of Teachers of Mathematics 2000) even while modeling activities remain on the fringe in many classrooms (Windschitl, Thompson, and Braaten 2008). In this article, we contribute to existing research on students' model-based reasoning with a fine-grained look at the interaction between two modeling resources: spatial reasoning and the body.

Students' capacities for spatial reasoning and embodied thinking are likely to impact how they engage in model-based reasoning. With respect to the former construct, sophisticated spatial reasoning capacities are not only cross-cultural and diverse (O'Meara and Pérez Báez 2011) but also predictors of future participation in STEM fields (Wai, Lubinski, and Benbow 2009). Models, as spatial depictions of target systems, call on the spatial reasoning skills of the scientists and students who interact with them. From a complementary standpoint, recent empirical and theoretical research in embodied cognition (Gibbs 2005) underscores that spatial reasoning occurs less in a cognitive vacuum than in a body tethered to small and large visual features of the landscape. On challenging spatial reasoning tasks, for example, participants routinely externalize or offload their spatial thinking onto the hands and arms in gesture (Schwartz and Black 1996). Even more importantly, from an educational standpoint, actions of the body are increasingly becoming recognized as shapers of thought (Goldin-Meadow and Beilock 2010) and markers of learning (Gerofsky 2010). Taken together, these findings suggest that the nexus between gesture and spatial reasoning is rich with potential for educators concerned with cultivating students' model-based reasoning.

In the following, we explore how students negotiate the tension in their gestures between enacting different viewpoints and maintaining a coherent display of space and time in the model. For example, when modeling planetary orbit with gesture (Crowder, 1996), students might spin in circles on both feet (depicting the planet from the *character viewpoint*) or they might stand still while using the finger to trace several circles in the air (depicting an *observer viewpoint* outline of the planet). As students transfer from one viewpoint to another for different components in the system, embodying at various points planets, suns, and moons at different scales and locations, students are still tasked with assembling a model that maintains some geographical and temporal consistency. In the space in front of and around the body, a student may gesture a fist-shaped planet located to the left, to the right, or in the center, moving speedily across or drifting lethargically. Because gestures occur in a topology of space and time, they present visible and tangible landscapes throughout which speakers can relate the locations and movements of model components. How students drift between different gestural viewpoints while still maintaining spatiotemporal continuity in the model—and what implications this has for learning outcomes—are our core concerns in this chapter.

We address the above research questions with a methodology in which students gesture in response to a sequence of verbal prompts, a protocol that elicits what we call *listening gestures*. The instruction to gesture in response to verbal prompts allows students to employ the body as a vehicle for organizing new information, while at the same time generating a powerful real-time display of the stages of comprehension for the researcher. In this way, our research design recruits gesture as a probe to investigate the learning process, analogous to a learning-aloud protocol (Clement and Steinberg 2002), but with a heightened spatiotemporal immediacy. We select for our analysis two students who responded very differently to this activity, each in her own way typical of the bi-modal distribution in our full study[1], and each illuminating the microgenetic change in the learning process.

2 Inanimate character and observer viewpoints

One of the modeling resources we examine in this paper, viewpoint, describes how students move their bodies *as* the component in the model or *as an observer* of the component in the model. Numerous studies of STEM environments have meaningfully documented how students' and teachers' adoption of multiple and flexible viewpoints impacts teaching and learning (de Freitas and Sinclair 2012; Crowder 1996; Enyedy et al. 2012; Gerofsky 2010; Hall 1996; Lindgren 2012; Nemirovsky and Monk 2000; Ochs et al. 1996; Warren et al. 2001; Wilensky and Reisman 2006). To further educational researchers' understanding of viewpoint, and to set the stage for its involvement in learning, we begin by reviewing research on gestural viewpoint from cognitive linguistics' research on storytelling.

The spoken words of stories often sound out to the accompaniment of gestures, and these gestures often show the actions of story events from different viewpoints. For example, in telling the story of Babe Ruth's legendary Called Shot, a speaker might use his own body to show how Babe Ruth pointed toward the outfield bleachers. The speaker might then *step out* of Babe Ruth's body and trace an arc in the air showing the flight of the home run ball clearing the outfield fence. These two gestures instantiate from different viewpoints: character viewpoint (C-VPT) and observer viewpoint (O-VPT) (McNeill 1992; Parrill 2009). The character viewpoint refers to gestures that reenact the action of the story character such that the gesturer and the story character move their bodies in similar ways (McNeill 1992; Parrill 2009). In the C-VPT Called Shot example, the storyteller moves part of his own body in the same way Babe Ruth moved his body in the past; both bodies point toward a distant object. On the other hand, observer viewpoint gestures involve the body conforming to how an object or character appears when observed from a distance. In the O-VPT Called Shot example, the storyteller uses part of his own body to display the trajectory of the baseball as seen from afar.

In a departure from the stories traditionally investigated in cognitive linguistics, we study STEM models in which the components lack human-

---

1 Our research team is still analyzing the data from the full study. We would like to acknowledge the support of the UCLA Office of the Vice Chancellor for Research and the Academic Senate Council on Research for Transdisciplinary Seed Grant A1118 and the National Science Foundation Grant 1028381. This material is also based upon work supported by the National Science Foundation Graduate Research Fellowship under Grant No. DGE-0707424.

like bodies, such as planets, computer data, or atoms. However, many of the character viewpoint definitions in the field take for granted an approximate one-to-one mapping between the bodies of the story characters and the body of the storyteller. Maws of cartoon dogs map onto the mouth of the storyteller and arms of superheroes onto human arms. Some definitions of C-VPT even exclude objects that cannot have viewpoints, such as baseballs, buildings, or clouds (Parrill 2009), classifying those gestures instead as metaphoric (McNeill 1992). Accordingly, the character viewpoint gesture, in its traditional sense, can be defined by the storyteller and the story character enacting actions with homologous articulators (e.g. if Sylvester the Cat runs in the cartoon on his feet, the gesturer telling the story about the cartoon simulates running also on his own feet); the storyteller uses his own 'articulator to represent a character's articulator' (Parrill 2012: 104). These prototypical targets are well studied, but we should be careful not to underestimate the creativity or fertility of the character and observer viewpoints, which extend beyond these natural mappings (Crowder 1996; Lindgren 2012; Ochs, Gonzalez, and Jacoby 1996).

In structuring gesture enactments metaphorically (Parrill and Sweetser 2004), speakers can depict inanimate components (such as electrons, computer data, or planets) from both character and observer viewpoints. In an education context, it is important to classify metaphorical gestures (McNeill 1992) along the dimension of viewpoint because the viewpoint dynamically structures how the model unfolds in space and time. As students need to reason carefully about spatiotemporal relations to understand new models, educational researchers should be attentive to how space and time interact with gestural viewpoint, even in the context of models of inanimate material. In this chapter, we attempt to track how inanimate C-VPTs and inanimate O-VPTs shape students' reasoning about spatial and temporal characteristics of models during a learning activity.

3 Spatiotemporal topology

While the distinction between character and observer viewpoint addresses the mappings between the body of the speaker and the features of the target modeled, the spatiotemporal topology refers to the scale of the gestured components, the location throughout which those components move, and the structure of events in time. Students exploring a model need to develop an understanding not only of how the shape of their bodies map onto objects in the model, but also of how the objects are positioned or move relative to each other over time.

These spatial aspects of gesture have been extensively documented in the literature. Descriptions of spatial concepts are frequently accompanied by co-speech gestures (Alibali, Heath, and Myers 2001). For example, college students perform gestures while reasoning about spatial puzzles (Emmorey and Casey 2001) and mechanical models (Schwartz and Black 1996), and adults and children perform gestures while talking about solutions to the spatial Tower of Hanoi puzzle (Garber and Goldin-Meadow 2002). As problem solvers animate the components of spatial puzzles with gestures, they make visible the spatial characteristics of the model. That is, if the hand gesture revolves counter-clockwise, the puzzle piece to which the gesture refers is also understood to revolve counter-clockwise (Emmorey and Casey 2001). In addition, an observer viewpoint gesture can be presented as spatially close to the story action (*inside* or *proximal perspective)* or far away from the story action (*outside* or *distal perspective)* (Crowder 1996; Gerofsky 2010; McNeill 1992). This body of work suggests that a variety of spatiotemporal topologies play an integral role in co-speech gesture.

In this study, we focus on two constructs within the area of spatial reasoning: scale and topology. Scale refers to the depiction of the size of the component. ASL signers, for example, can represent the same component at different scales on subsequent turns. A signer might depict a motorcycle at the scale of the full human body in one turn and then depict the same motorcycle at the scale of a small object that fits in the palm of the hand on the next turn (Dudis 2004). The motorcycle shrinks in size from an object that a person can ride to something you can hold in your hand. Topology tracks how gestured elements create an organized geographical layout of objects, even if transient. While talking about family lineage, for example, Lao speakers make spatially consistent points toward empty space as if family tree diagrams were floating in front of their bodies. As a cognitive artifact, the body becomes 'a visuospatial representational resource' (Enfield 2001: 1). Environmentally coupled gestures (Goodwin 2007)—gestures that reference physical objects in the space around the body—are grounded in the given spatial layout of structures in the environment, yet uncoupled gestures must also conserve a spatial topology over some time window, however brief. A fascinating literature explores cultural variability in the spatial language and spatial thinking that occurs (in broad strokes) either relative to the speaker's body or relative to objects or geographical landmarks present in the environment (for an overview, see O'Meara and Pérez Báez 2011).

These constructs point to the need to track how students organize the scale of gestured representations and the spatial arrangement of those representations as they learn to model a new system. By examining how students establish and manipulate the spatiotemporal topology of the modeled system, in conjunction with their adoption of multiple viewpoints, we aim to illuminate a critical dimension of learning.

4 Integration across viewpoint and space: A case study

In bringing together the constructs of gesture viewpoint (observer or character) and gesture space and time (spatiotemporal topology), we find a potential challenge to the process of conceptual integration. Modeling requires an understanding of both the repertoire of actions that components can take and the way those actions relate to other components (Lehrer and Schauble 2006). It is not enough to know that planets rotate; the student also needs to know *where* planets rotate and *around which* other planets. In the gesture stream, the student can both feel the action of the component and track how that action relates geographically to previous and future actions. The body provides the former feeling equipment and the empty area around the body provides the later location for organizing spatiotemporal landscapes.

The multi-viewpointed nature of cognition poses a challenge to learners: How do students dynamically transition between character and observer viewpoints while maintaining a sense of how events in one part of the system relate to events in other parts of the system? The switch from viewpoint to viewpoint in the model may demand re-presentations of scale and/or space. While research suggests that sensory metaphors are rife in human language and recruited in a loosely organized 'in-pieces' fashion (diSessa 1993), we know less about how students coordinate multimodal knowledge fragments into coherent wholes. The conceptual blending framework (Fauconnier and Turner 1998) presents a cognitive mechanism, but empirical approaches are needed to detail the process of conceptual integration during learning activities.

5 Method

The present case study was part of a larger study conducted at a West Coast University with college students (*N*=20) aged 18-22. Restrictions based on demographics were not used, but participants with extensive existing knowledge of the target concept, packet switching, were excluded from the study. The study used convenience sampling of undergraduates enrolled in an introductory communications course and followed IRB guidelines.

Student participants learned the concept of packet switching, the logic of a digital network at the heart of the Internet. In packet switching, individual pieces of data (packets) move according to network rules between different relay nodes (routers). A message is broken up into packets at the sending computer, tagged with information about its origin and destination, and sent through a web-like network toward the receiving computer (see Figure 1). Copies are made at each router and used in the case of failed transmissions; receipts are returned to confirm delivery. Students were taught the concept of unicasting, which involves a chain of packets directed toward one destination, and students were also taught the concept of multicasting, which involves packets proliferating throughout the network in response to requests from any number of destination computers. In the activity instructions, we instructed half of the sample (*n*=10) to listen to verbal statements about the packet switching system and try to model the system with their bodies. As these gestures signify an interlocutor's words, we refer to them as listening gestures. The other half of the sample (*n*=10) listened to similar instructions but did not gesture. In this chapter, we report on only two students from the listening gesture condition.

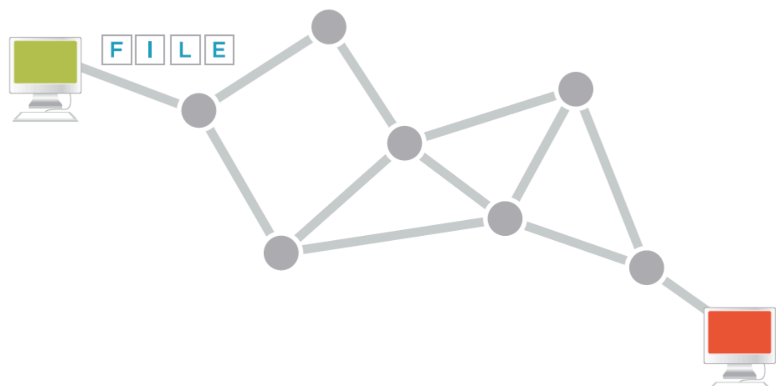

Figure 1. A packet-switching network, in which the message, FILE, is broken up into smaller packets—F, I, L, E—and routed through the network.

The activity took place in one-on-one interactions between the student and the researcher. Students were instructed to stand up and model each sentence with gestures, and informed that the researcher would read the full 27-sentence narrative twice (study phase 1 and 2). Each verbal prompt described a single, goal-direction action.sThe listening gestures protocol was designed on analogy with the learning-aloud protocol (Clement and Steinberg 2002) to gain insight into the processes of learning through real-time monitoring. The protocol is particularly useful for investigating the role of spatiotemporal representations in concept development, contributing to the understanding of the multimodal dimensions of human cognition and communication (Steen and Turner, this volume). As a mandatory learning strategy, the protocol is disruptive, typically either facilitating or hindering learning and potentially useful for contrastive studies. The factors affecting the nature of the disruption were not investigated in the current study.

The study phase of the session was video recorded with a stationary camera placed in the corner of the room and focused on the participant. The narrative describing packet switching contained verbs typically used to describe human actions, such as grab, travel, and cut. These verbs metaphorically structure the actions of packets and routers through human actions, for example, as in the phrases, '…it chops up the email' and '…before throwing the packet toward the next router.' Each prompt was read with clear enunciation (on average 3 words per second) with a 4-second pause between each prompt. The experimenter produced no gestures and participants did not ask any questions during the reading. After the study phase, students walked across campus to an fMRI facility where they responded aloud to 45 open-ended questions about packet switching from inside the scanner (the fMRI part of the study will not be reported here). After walking back to the lab, the researcher asked the student to explain in words everything he/she remembered about the operation of packet switching, what we call the *explanation phase*. The explanation phase was meant to evoke a round of unprompted co-speech gestures. Finally, students were given a written post-test with open-ended questions about packet switching.

5.1 Measures

Character versus observer viewpoint: Real packets and routers on the Internet do not, of course, have hands, feet, and a head. However, we propose that the definition of an inanimate character viewpoint can build on Parrill's (2012) C-VPT definition: the speaker and the story character conduct an action with homologous body parts. In the inanimate context, the gesturer construes the inanimate component as having a human-like articulator. The storyteller can map his own body part onto a figurative but analogous hand, finger, body, or head of the inanimate component. That is, the gesturer can metaphorically construe routers as having *hand-, feet-,* or *head-like* articulators. Whenever the inanimate component is construed in the gesture as having a human-like articulator (e.g. a router has a *hand*) and the gesturer uses his/her own analogous articulator (e.g. his own hand) to enact an action of the component (e.g. throwing), we consider that the gesturer has adopted the character viewpoint of that component. Examples would include a packet with a hand that can *grasp*, a packet with a hand that can *scribble*, a router with a head that can *look*, or a packet with arms that allow it to *run*. Alternatively, an observer viewpoint (O-VPT) gesture would constitute representations that depict the shape or trajectory of a component. The hand plays the role of an outsider sketching descriptions of component entities. The target component, in this case, is not construed as having an homologous human articulator, but rather, the articulator of the gesturer forms the shape or trajectory of the component. Examples include a closed fist that represents a motionless packet, a flat hand that depicts the trajectory of a moving packet, or both hands tracing the square shape of a router.

Spatiotemporal topology: As the gesture-based simulation of the system unfolds, the gestural representations offer a visual landscape that details the locations and sizes of components in packet switching. The two dimensions of scale and location comprise a partial description of the spatiotemporal topology. Take the example of a gesture in which the fist represents a small packet moving from right to left across the body. The fist represents the size; the right to left (transverse) axis on which the packet moves represents direction of movement. Gestural models, in this way, provide a rich window onto the spatial organization of representations. As participants formulate listening gestures in response to words describing packet switching, they can interpret verbal spatial descriptions relative to either the past locations of gestured components or the present location of the body. In the former object-centered frame of reference, spatial features of sentences are encoded with respect to the direction of movement of components from previous gestures. In the later relative frame of reference, the body of the gesturer encodes the spatial meaning (O'Meara and Pérez Báez 2011). For example, object-centered spatial reasoning would map the word *left* onto the left part of the axis on which a gestured packet moved in the previous turn. Relative spatial reasoning would map the word *left* onto the left of the gesturer's body regardless of where the gestured packet just moved. Throughout this text, the terminology for spatial relations will use the body of the gesturer as the reference point (e.g. right refers to the right of the gesturer's body from the gesturer's viewpoint).

6 Results

In this section, we detail how two participants—Jen, the student pictured with light hair, and Beth, the student pictured with dark hair (names are pseudonyms)—gestured a subset of packet switching concepts during study phases one and two and during the explanation phase. What struck us was how these two students differed in their approaches to organizing representations in space. We selected Jen and Beth as prototypical members of two clusters that evenly divided the sample. Our research team is currently analyzing the results from the full sample.

6.1 Jen on source and destination computers and the packet trajectory

Row 1 of Figure 1 contains snapshots of Jen in study phase one creating her first listening gestures that model the start and destination of the packet journey (S1a, S1b) and the pathway the packet takes (S1c). Jen then gestures these concepts a second time in round two of the study phase (S2a, S2b, S2c) and a third time as co-speech gestures accompanying her own description in the explanation phase (E1a, E1b, E1c).

In round one, Jen gestures first in response to the prompt, 'the computer stamps each packet with a precise tag that tracks three things: the starting point of the journey, the destination of the journey … ' (S1a, S1b), and then in response to the prompt, 'the packets rush out from the computer into the network' (S1c). With respect to the first prompt, Jen points in sequence toward the space to the left (S1a) and to the right (S1b) of her body, indicating the location of the source computer and the destination computer. For the second prompt, she traces with both fingers the trajectory of packets entering the network. In this turn, Jen starts at the center of her torso and traces a loop with each hand toward the sides of her body. Then, Jen returns her hands to the center of her torso and traces two more loops in the direction directly out in front of her body (S1c).

These gestures generate a skillful display of the verbal description. The hands point to two empty spaces at opposite ends of the body to signify the start and end of the journey. Then, the hands dynamically trace the pathway that multiple packets take through the network, spreading out like arcs in a

Figure 2. Jen gestures the source and destination locations & the packet trajectory.

fountain. Without any training or any feedback, Jen maps the language onto her body, referring to locations and pathways from an observer viewpoint. In round two of the study phase, Jen constructs listening gestures for the same two concepts in response to the same two prompts. This time, the full hands stand in as the shape of the source (S2a) and destination (S2b) computers, and then instead of tracing a packet, Jen steps into the character viewpoint of the source computer and simulates the motion of throwing the packet into the network (S2c). All three gestures re-appear when Jen, over an hour later, explains in her own words her understanding of packet switching (E1a, E1b, E1c). She again uses her hands to stand in as the source and destination computer (while saying 'it goes from a single source to another single source'), and then simulates a computer throwing a packet into the network (while saying, 'the packet … is sent'). Agency is granted to the computer without which the packet would not enter the network. In round two and in the explanation phase, Jen shifts seamlessly from the viewpoint of an observer marking the location of two computers to the viewpoint of a character (the source computer) launching packets.

For each prompt, Jen creates a creative, original, and intelligible mapping of words onto her own body, using both character and observer viewpoints. As she gestures, she *simultaneously* builds up an organized spatiotemporal topology. That is, the source and destination computers always materialize in specific locations: the left of the body and the right of the body respectively (S1a, S1b, S2a, S2b, E1a, E1b). The packets also move along a constant sagittal axis starting at the center of the body and moving forward directly in front of the body (S1c, S2c, E1c), with a few exceptions to this rule in the explanation phase. Examined in this light, we notice that Jen creates a sequence of individual, disconnected spatiotemporal topologies. In the second prompt, the packet, upon entering the network, starts not at the location that Jen depicts the source computer in the previous prompt, but rather from a new location in the center of the body. In addition, instead of heading in the direction of the destination computer from the previous prompt, the packet moves on a new axis extending out in front of the body, headed toward an unknown destination. The pattern of placing source and destination computers to the left and right on the transverse axis and then sending packets from the center toward the front on the sagittal axis appears in study phase one and two and in the explanation phase.

In the transition from the first to the second prompt, then, Jen starts more or less afresh, rebuilding the spatial topology. She generates an intelligible representation of each prompt from a blank slate. The gestures are memorable, repeating again in the second round and in the explanation phase as they establish a fixed pattern. Importantly, the spatiotemporal topology is consistent *within* each of Jen's gestures for a single prompt. That is, the source and the destination appear in different locations and one after the other in time. Similarly, the packet moves in time from one place to another along a direct spatial route. It is *across* these prompts that the spatial topologies diverge. We have noted that in study phase round two and in the explanation phase, Jen transitions from an observer viewpoint showing two computer locations to a character viewpoint acting as the transmitting computer. In the former, the hands depict the computer shape; in the later, the source computer has a *hand*, and Jen shows that hand throwing material toward the network. The viewpoint shift triggers a scale shift. That is, in the former, packets would be the size of small specks on the hand. In the later, the packet would be the size of a tennis ball. Looked at another way, the throw from the computer viewpoint in the second prompt takes up about as much space as does the full network in the previous prompt. In short, the viewpoint switch results in a scale switch that may be partly responsible for driving Jen to rewrite the model's spatial order.

6.2 Beth on source and destination computers and the packet trajectory

Figure 2 shows Beth in a parallel set of snapshots to Jen in Figure 1. While listening to the first of the two prompts in study phase round one, Beth places emblematic symbols for the number one (a single index finger, S1d) and the number 3 (three fingers, S1e) in two different locations (left of the body and right of the body, respectively), showing the start and destination of the packet journey. The number symbols mark that the tag on the packet tracks three items, two of which are the source and destination locations. In this way, the location of the gestured number symbol corresponds to the location of the computers while the number itself refers to the number of items the packet tag stores. In the second round of the study phase, Beth more simply points to the two computer locations (S2d, S2e). In the explanation phase, Beth curves her hands around what appear to be two small objects (E1d, E1e) to depict the source and destination computers (while saying, 'you have a computer over here and that's the starting place and you have a computer over here and that's the ending place'). In this way, Beth symbolizes the start and finish in three different ways: locations of emblematic numbers, locations revealed through deictic gestures, and locations of iconic objects around which the hand wraps. As in Jen's case, the spatiotemporal order of these symbols is constant within each prompt: the source computer always appears first and to the left while the destination computer appears second and to the right.

Jen and Beth differ, however, in the degree to which they spatiotemporally integrate adjacent prompts. Both in rounds one and two of the study phase, Beth uses her full left hand to depict an O-VPT shape of a packet and

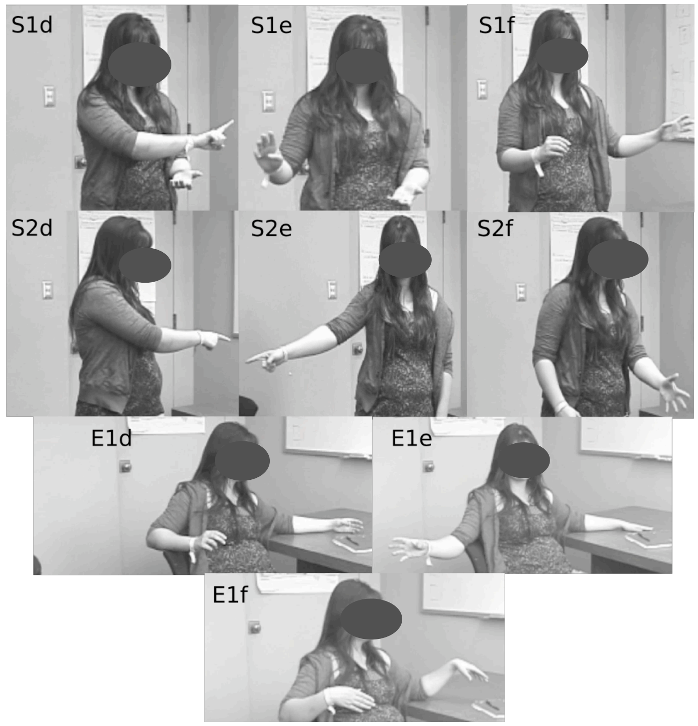

Figure 3. Beth gestures the source and destination locations & the packet trajectory.

shows that packet departing (S1f, S2f) from the exact location of the source computer in her model of the prior prompt. In addition, she shows the packet moving in the direction of the prior location of the destination computer. Accordingly, the spatiotemporal organization of the model in the first prompt—source computer first and to the left, destination computer second and to the right—guides how Beth spatiotemporally organizes the second prompt—packets move from left to right. In the explanation phase, even when Beth adopts the viewpoint of the source computer, modeling the act of throwing the packet (E1f) into the network (while saying, '…those get sent out'), she starts the throw from the prior location of the source computer.

Beth's listening gestures modeling the second of the prompts are guided by the spatiotemporal topology that she created in the model of the previous prompt. She links the organization of time and space across prompts, in effect generating a single underlying spatiotemporal representation. The spatiotemporal order exists despite Beth using all three of emblematic, deictic, and iconic gestures and even switching from an observer viewpoint detailing packet shape to a character viewpoint symbolizing a computer throw.

Furthermore, the spatiotemporal topology is conserved across both rounds of the study phase and even an hour later during the co-speech gestures of the explanation phase.

6.3 Multicasting on two pathways

In this section, we detail another example of Jen providing within-prompt spatiotemporal integration and of Beth providing both within-prompt and across-prompt integration. In this part of the study phase, the researcher reads aloud the following two prompts describing multicasting: ‘When the packet reaches the router, it reveals its tag. In this case, two different destination computers need the packet: one located at the end of the left pathway leading away from the router, and one at the end of the right pathway leading away from the router.’

In responding to the first of these prompts, Jen simulates grabbing a small bit of material from the space in the center of her body, carries it in her right hand to the right, and then opens the hand. In response to the second prompt, Jen makes an emblematic gesture with two fingers to show the number two, and then presents an upward-facing, left-handed flat palm to the left of her body followed by an upward-facing, right-handed flat palm to the right of her body. The C-VPT action of grabbing and transferring the packet in the first prompt occurs along a transverse axis from center toward right space. The O-VPT depiction of the location of the destination occurs in space to the left of the body and to the right of the body. In this way, the spatial organization of the gestures is not conserved across prompts. Packets move from left to right on the transverse axis and then from the center out in front on the sagittal axis. In summary, for the second half of the prompt, instead of construing the words *left* and *right* from the packet viewpoint (an object-centered frame of reference), Jen depicts the words *left* and *right* with respect to her own forward-facing body (a relative frame of reference). Within each prompt, the gestural model is spatially organized; across prompts, the layouts diverge.

Beth, on the other hand, does organize the spatial layout of the model across prompts. In response to the first prompt, she traces an O-VPT trajectory of a packet moving from left space to center space, before modeling in center space the C-VPT packet prying open its own tag with both hands. For the second prompt, Beth indicates the left and right pathway toward the two destination computers relative to the existing axis and direction on which the packet just traveled (an object-centered frame of reference). That is, Beth traces two different pathways to the space to the right of her body, one extending slightly to the left and the other slightly to the right of the packet’s earlier transverse trajectory. In this way, Beth presents a packet that never veers from the transverse axis on which it travels. The packet moves along the axis from left to center, headed toward destination computers to the right. Each new entity or event described in words appears in a logical spot on this axis, even when Beth moves from an O-VPT packet depiction to a C-VPT packet depiction and then back to an O-VPT pathway depiction. In round two of the study phase and in the explanation phase, Beth creates the same spatial topology. Furthermore, in round 2, she moves her full body into the space to the right, turns back toward center, and makes a pulling 'come here' motion, showing from a C-VPT the destination computers requesting the packet. The character viewpoint request gesture, then, is shifted over to the prior spatial location of the destination computer. This effectively laminates a relative spatial framework onto an object-centered framework. Instead of thinking about how destination computers *somewhere out there* would construe space, Beth moves her whole body into the location where destination computers had appeared in her past gesture layouts. Beth can then use her own relative space as synonymous with the computer's object-centered space.

6.4 Learning comprehension

How does spatiotemporally constructed conceptual integration affect learning comprehension? As the results we present here are based on two participants, we have no basis for inferring causation or even correlation, yet this case study suggests the need for follow-up research. Jen, although near perfect in depicting the meaning of each individual prompt, disrupts the spatiotemporal topology in round one on 46 percent of her gestures and in round two on 30 percent of her gestures. Beth, on the other hand, disrupts the spatiotemporal topology in round one on 6 percent of her gestures and in round two on 3 percent of her gestures. That is, both students correctly interpret the actions described in each prompt, but Beth goes to greater lengths to fit those actions into a highly organized spatiotemporal topology. Beth's gestures are coherent with respect to the topology expressed in the previous gesture, creating a cumulative topology; her gestures provide evidence of the formation of a coherent and tacit underlying spatiotemporal representation.

On the post-test, Jen correctly describes many of the actions that take place in packet switching: packets receive tags, routers receive receipts, routers make copies. However, Jen struggles to draw correct inferences about questions that require stringing together multiple actions in the system. She is unable to determine how many copies would be made when packets pass through multiple routers in unicasts and in multicasts, where and when those copies are made, and how routers know where to send packets. The actions that comprise these topics are individually described in the prompts, but they must be systematically related to other prompts to reveal the answers. In the explanation phase, Jen leaves out almost every step of the multicasting process and conflates aspects of unicasting and multicasting. Beth, on the other hand, is able to recall individual actions in packet switching and draw inferences across actions. She correctly reasons about the number of copies made in unicasting and in multicasting and the locations where those copies are made. In the explanation phase, she correctly describes all of the aspects of unicasting and multicasting.

7 Discussion

Complex models must typically be taught by breaking them down into simpler components. Each component must be understood separately, a task that may require both creativity and dedication. Yet this is not sufficient: the whole must be constructed out of a series of specific relations between the parts. A genuine understanding emerges out of a dynamic integration of components that preserves their known causal relations. This secondary integration cannot be taken for granted.

In this study, using listening gestures as an investigative activity, we induce and uncover a contrastive dynamic in a two-stage process of conceptual integration. Each verbal prompt was carefully designed to generate an individual gesture event, defined as a single, goal-directed action. In describing our participants' gestures, we note how the adoption of a particular character or observer viewpoint generates a distinctive dynamic topology, in which each element in the prompt is modeled as a rich spatiotemporal process. In the first stage, event-based causal relations are modeled as bodily motions in space in ways that are highly metaphorical; for instance, copying an electrically encoded packet of information is modeled by a gestural metaphor of grasping a fist-sized object and moving it to an adjacent location. A complex conceptual integration network is required to project the relevant features of the spatiotemporal topology onto domain-specific content such as packets, routers, and cables from the target, resulting in a creative blend that constitutes a preliminary understanding of an action from a single prompt.

We see that in Jen's case, each individual model is created afresh, with no relation to the preceding. This modular strategy results in a basic understanding of the target model, along with limited abilities to generate new inferences and a failure to grasp the more complex behaviors of the system. In contrast, Beth adopts an integrative strategy that results in a richer understanding of the target model, associated with enhanced abilities to generate new inferences and a confident grasp of the more complex behaviors of the system.

What is Beth doing? Our analysis suggests she uses the movement of her body in space and time to build an integrated spatiotemporal model, a learning strategy we call thinking with the body. In response to each verbal prompt, she adopts a suitable character or observer viewpoint and generates a distinctive gesture to model a particular component of the target system in space and time. This spatiotemporal topology in turn orients and guides her subsequent gesture, which may adopt a different viewpoint, develop a new gestural metaphor, and generate a new and supplementary extension of the spatiotemporal model. By attending to her own creations as a context at each stage for her new creations, she cumulatively builds a model that encodes not only the individual components in an accurate manner, but also their internal relations (Figure 3).

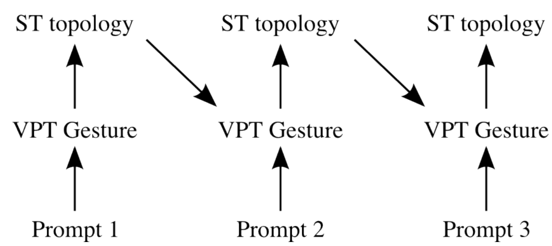

Figure 3. Thinking with the body involves a recursive process of spatiotemporal continuity linking gestural events across changing viewpoints.

What we have aimed to show in this chapter is that gestural acts, in generating complex patterns in space and time, provide a dynamic scaffolding for conceptual development. Despite that the conceptual integration networks for individual gestural acts have been extensively studied, the application of gestural studies to the Learning Sciences requires that we examine the integration of multiple gestural acts into complex and dynamic representations. Through two contrasting case studies, we have shown how a succession of viewpointed gestural acts in a recursive manner can become incorporated into and simultaneously contribute to the elaboration of a cumulatively constructed spatiotemporal topology, possibly facilitating high-level modeling with an impressive power to generate new inferences.

References

Alibali, M. W., D. C. Heath, and H. J. Myers 2001. Effects of visibility between speaker and listener on gesture production: Some gestures are meant to be seen. *Journal of Memory & Language* 44:169-188.

Clement, J. J. and M. S. Steinberg 2002. Step-Wise Evolution of Mental Models of Electric Circuits: A "Learning-Aloud" Case Study. *The Journal of the Learning Sciences* 11(4):389-452.

Crowder, E. M. 1996. Gestures at work in sense-making science talk. *The Journal of the Learning Sciences* 5(3):173-208.

de Freitas, E. and N. Sinclair 2012. Diagram, gesture, agency: theorizing embodiment in the mathematics classroom. *Educational Studies in Mathematic*s 80:133-152.

diSessa, A. A. 1993. Toward an epistemology of physics. *Cognition and Instruction* 10:165- 255.

Dudis, P. G. 2004. Body partitioning and real-space blends. *Cognitive Linguistics* 15(2):223- 238.

Emmorey, K. and S. Casey 2001. Gesture, thought and spatial language. *Gesture* 1:35–50.

Enfield, N. J. 2005. The body as a cognitive artifact in kinship representations: Hand gesture diagrams by speakers of Lao. *Current Anthropology* 41(6):51-81.

Enyedy, N., J. A. Danish, G. Delacruz, and M. Kumar 2012. Learning physics through play in an augmented reality environment. *International Journal of Computer Supported Collaborative Learning* 7(3):347-378.

Fauconnier, G. and M. Turner 1998. Conceptual integration networks. *Cognitive Science* 22:133-187.

Garber, P. and S. Goldin-Meadow 2002. Gesture offers insight into problem solving in adults and children. *Cognitive Science* 26:817–831.

Gerofsky, S. 2010. Mathematical learning and gesture: Character viewpoint and observer viewpoint in students' gestured graphs of functions. *Gesture* 10(2-3):321-343.

Gibbs R. W. Jr. 2006. *Embodiment and Cognitive Science*. New York: Cambridge Univ. Press

Gilbert, J. K. 2000. *Developing models in science education*. Dordrecht, The Netherlands: Kluwer.

Goldin-Meadow, S. and S. Beilock 2010. Action's influence on thought: The case of gesture. *Perspect Psychol Sci* 5(6):664–674.

Goodwin, C. 2007. Environmentally coupled gestures. *Gesture and the Dynamic Dimension of Language*, ed. S. D. Duncan, J. Cassell and E. T. Levy, 195-212. Amsterdam: John Benjamins.

Hall, R. 1996. Representation as shared activity: Situated cognition and Dewey's cartography of experience. *Journal of the Learning Sciences* 5(3):209–238.

Lehrer, R. and L. Schauble 2006. Cultivating model-based reasoning in science education. *The Cambridge handbook of the learning sciences*, ed. R. K. Sawyer, 371-388. New York: Cambridge University Press.

Lindgren, R. 2012. Generating a learning stance through perspective-taking in a virtual environment. *Computers in Human Behavior* 28(4):1130-1139.

McNeill, D. 1992. *Hand and mind: what gestures reveal about thought*. Chicago: University of Chicago Press.

National Council of Teachers of Mathematics (NCTM) 2000. *Principles and Standards for School Mathematics*. Reston, VA.

National Research Council 2002. Scientific research in education. *Committee on Scientific Principles for Educational Research*, ed. R. J. Shavelson and L. Towne, Washington, DC: National Academy Press.

Nemirovsky, R., C. Tierney, and T. Wright 1998. Body motion and graphing. *Cognition and Instruction* 16(2):119-172.

Ochs, E., P. Gonzalez, and S. Jacoby 1996. When I come down, I'm in a domain state: Grammar and graphic representation in the interpretive activity of physics. *Interaction and grammar*, ed. E. Ochs, E. A. Schegloff, and S. Thompson, 328-369. Cambridge: Cambridge University Press.

O'Meara, C. and G. Pérez Báez 2011. Spatial frames of reference in Mesoamerican languages. *Lang. Sci*. 33:837–852.

Parrill, F. and E. Sweetser 2004. What we mean by meaning: Conceptual integration in gesture analysis and transcription. *Gesture* 4(2):197-219.

Parrill, F. 2009. Dual viewpoint gestures. *Gesture* 9(3):271-289.

2012. Interactions between discourse status and viewpoint in co-speech gesture. *Viewpoint in Language: A Multimodal Perspective*, ed. B. Dancygier and E. Sweetser, 97-112. Cambridge: Cambridge University Press.

Schwartz, D. L. and J. B. Black, J. B. 1996. Shuttling between depictive models and abstract rules: Induction and fall-back. *Cognitive Science* 20:457-497.

Schwarz, C. V., B. J. Reiser, E. A. Davis, L. Kenyon, A. Acher, D. Fortus. . . J. Krajcik 2009. Developing a learning progression for scientific modeling: Making scientific modeling accessible and meaningful for learners. *Journal of research in science teaching* 46(6):632-654.

Wai, J., D. Lubinski, and C. P. Benbow 2009. Spatial Ability for STEM Domains: Aligning Over 50 Years of Cumulative Psychological Knowledge Solidifies Its Importance. *Journal of Educational Psychology* 101(4):817-835.

Warren, B., Ballenger, C., Ognonowski, M., Rosebery, A. S., and Hudicourt-Barnes, J. 2000. Rethinking diversity in learning science: the logic of everyday sense-making. *Journal of research in science teaching* 30(5):529-552.

Wilensky, U. and K. Reisman 2006. Thinking like a wolf, sheep, or a firefly: Learning biology through constructing and testing computational theories--An embodied modeling approach. *Cognition and Instruction* 24(2):171-209.

Windschitl, M., J. Thompson, and M. Braaten 2008. Beyond the scientific method: Model-based inquiry as a new paradigm of preference for school science investigations. *Science Education* 92(5): 941–967.