15. Cognitive Anthropology: Distributed cognition and gesture

Abstract

All human cognition is distributed. It is, of course, distributed across networks of neurons in different areas of the brain, but also frequently across internal and external representational media, across participants in interaction, and across multiple spans of time. While most research on distributed cognition has focused on complex activities in technology-laden settings, the principles apply just as well to everyday cognitive activities. Studies of distributed cognition reveal that bodily activity — especially gesture — plays a central role in coordinating the functional systems through which cognitive work gets accomplished. Gesture does more than externalize thought; it is often part of the cognitive process itself. Gestures create representations in the air, enact representational states on and over other media, and bring states in different media into coordination to produce functional outcomes. Gesture also plays a central role in propagating functional systems — associated cultural practices, cognitive models, and forms of coordination — across generations, while adapting them to the particulars of new problem situations. In so doing, gesture helps to sustain and enhance the cognitive sophistication of the human species.

1. Introduction

A child told to share candies with her sibling doles them out one at a time, saying "one (for you), one (for me); two, two; three, three…." Or she spreads them across a table and points at candies in succession while saying "one, two, three…." Or her mother shows her how to slide candies across the table two at a time while counting "two, four, six…." Each scenario involves hand actions, one pointing, and one demonstrating. Are these actions gestures?

What we call gestures and how we study them reflect particular theories of human cognition and communication. The traditional view in cognitive science is that humans think internally through propositional logic and/or mental imagery and express their thoughts to others through language, spoken or written. Spoken messages are accompanied by paralinguistic cues, such as vocal tone, facial expression, or body language, that signal emotional state or stance toward what is being said.

Head movements signal agreement or disagreement, and hands support or supplement the content of spoken messages. These signals help listeners unpack utterances and recover their propositional and/or imagistic content. From this point of view, communicating is a matter of encoding and decoding messages transmitted from sender to receiver: what since Reddy (1979) has been called the "conduit metaphor" of communication. This view of cognition and communication served as the basis for most research in cognitive science during the mid-1950s to the late 20th century, until other views, including that of distributed cognition, called this account increasingly into question

Against this backdrop, gesture re-emerged as a topic of research due primarily to the pioneering studies of Adam Kendon (1972, 1980) and David McNeill (1985, 1992). Kendon and McNeill both study the expressive hand movements that accompany speech, which Kendon (1980) calls "gesticulation," and how they relate to spoken content. For both researchers, a primary unit of analysis is the utterance, a communicative act consisting of a speech-gesture ensemble expressing related content and bounded as a single intonation unit (as in Chafe 1994). For Kendon (2004), speech and gesture are separate streams coordinated in the process of utterance, while for McNeill (2005), speech and gesture arise together, in a dialectic of language and imagery, from a single idea or "growth point." Gesture, like speech, is a means for expressing or externalizing thought in the mind of the speaker, although it can also mark or regulate aspects of the discourse. Kendon's data consist primarily of recordings of conversations and some guided tours, while McNeill's consist mostly of experimental participants narrating events seen in a cartoon or film or recalled from a fairy tale. The gestures examined in these studies are produced in the air in the space in front of the speaker or, in the case of pointing gestures, directed toward objects or locations in the surrounding space. The ground-breaking studies of Kendon and McNeill contrast in some respects with the workplace studies typical of distributed cognition research, where gestures over representational artifacts are common and where gesture and speech are directed toward the accomplishment of joint activity as well as the development of mutual understanding.

Studies of distributed cognition are closer to the work of Jürgen Streeck (2009) and Charles Goodwin (2000), researchers from the tradition of conversation analysis who study the communicative practices of people engaged in work and life activities in the culturally rich settings they ordinarily inhabit. These researchers study gesture as practice — as part of what people do and how they go about doing it — rather than as expressions of interior mental life. They take a particular interest in times when people coordinate with one another to develop a shared understanding of a problematic situation or to overcome snags in the flow of activity; here gesture comes to the fore. In these studies, gesture is entwined with practical action, so that gestures are frequently produced on or over objects in an "environmentally coupled" way (Goodwin 2007) or with objects in hand (as in LeBaron and Streeck 2000). Gesture may not be singled out for study but may instead be considered one of many factors shaping the construction of meaning *in situ*, including the structure of the activity, aspects of the setting, the position and orientation of participants' bodies (including access to each other's actions), mutual orientation toward objects, shared knowledge or history of activity, content and structure of the preceding discourse, and, of course, the talk that participants produce (Goodwin 2000). Attention is paid to conversational moves of various kinds (even inaction), and meaning is seen as emergent from the relations between talk, gesture, artifacts, and situated aspects of the discourse rather than from the unpacking of utterances.

Along with a focus on practice, distributed cognition research shares with these studies the use of micro-ethnography as a method of inquiry. Data consist of recordings of activity in real-world settings where the researcher is a participant-observer. Episodes of recorded activity are analyzed in detail — moment by moment, frame by frame — to reveal the subtle processes of coordination through which activity is accomplished and through which participants jointly construct meaning. Interpretations of the video data are warranted by evidence gathered through traditional ethnographic inquiry. The form of any particular analysis depends on the research questions posed and the conceptual framework employed in the study. In distributed cognition research, this micro-ethnographic approach is known as "cognitive ethnography" (Hutchins 2003, Williams 2006). Its goal is "to study how cognitive activities are accomplished in real-world settings" (Hutchins 2003): what resources are brought to bear and how they are coordinated to produce targeted outcomes. In other words, cognitive ethnography is close study of the phenomena that cognitive scientists seek to explain: human cognition in natural activity. As an approach focused on close observation and micro-analysis, cognitive ethnography can be combined with other methods of inquiry to enhance the ecological validity of experiments or to inform the design of simulation studies (Hutchins 2003). Together, these approaches help us triangulate toward a better understanding of human cognition.

Cognitive ethnographic studies of distributed cognition show that bodily actions, including gesture, play a central role in real-world cognitive activity. This article reviews key tenets of distributed cognition, briefly describes the role of the body in distributed cognitive functional systems, and highlights the affordances of gesture as a representational medium. It then discusses findings from cognitive ethnographic studies that illustrate critical functions of the hands in human cognition: creating and coordinating representational states in functional systems and guiding conceptualization to propagate functional systems across generations. The article concludes with brief implications for the study of gesture as a unique and powerful human capability.

2. Distributed cognition

The term "distributed cognition" refers not to a type of cognition but to a perspective for understanding cognition generally. As described by its leading proponent, Edwin Hutchins, all human cognition is distributed. Some cognitive accomplishments rely solely on interactions among neural networks in diverse areas of the brain, while others, including the most significant human accomplishments, involve coordination of internal structures and processes with structures in the world we engage with our bodies and modify to suit our purposes. Through such functional couplings, we use our Stone Age brains to lead Space Age lives.

Among the chief insights of distributed cognition is the benefit to be gained by not defining the boundaries of the cognitive system too narrowly. If we consider cognitive processes of reasoning, decision-making, and problem-solving to be those "that involve the propagation and transformation of representations" (Hutchins 2001: 2068; see also Hutchins 1995a: 49), then we must also consider that many of the most important representations lie outside the heads of individuals, embedded in sociotechnical systems of human activity. By incorporating relevant aspects of the material setting and social organization into the unit of analysis, we can study how cognitive systems function through "the propagation of representational state across a series of representational media" and how representational states are propagated "by bringing the states of the media into coordination with one another" (Hutchins 1995a: 117). Once we have an understanding of how such a distributed system functions, we are then in a position to make claims about what must be happening inside the heads of individuals to make the system function. Working from the outside in, we can "[refine] a functional specification for the human cognitive system" (Hutchins 1995a: 371) while avoiding the danger of over-attributing internal structure, that is, of claiming that more of the world is internally represented than is necessary to support adaptive behavior.

With respect to the field of cognitive science, this view of distributed cognition retains a sense of cognition as computation while dispensing with the notion of cognition as internal symbol processing. Cognition is foremost active, engaged, and embodied. Although it can play out covertly in imagined perception and action, the remarkable human capacity to ponder derives from a history of bodily engagement with the world. Human cognition is also, to a vastly greater extent than in other species, a mixture of the biological and cultural. Despite limited capacities for attention, memory, perception, and processing, humans are capable of stunning achievements primarily because "cultural practices orchestrate the coordination of low-level perceptual and motor processes with cultural materials to produce particular higher-level cognitive processes" (Hutchins 2010: 434). Distributed cognition views culture as, among other things, "an adaptive process that accumulates partial solutions to frequently encountered problems" (Hutchins 1995a: 354). These partial solutions structure systems of activity in which humans engage and through which they develop. Bodily actions, including gestures, bring these functional systems into coordination and perpetuate them across generations. As they do so, the systems adapt to changes in the cognitive ecology: to different environments, emerging technologies, new forms of social organization, and changes in cultural values and practices.

3. Using the body to coordinate elements in a functional system

While most research on distributed cognition has examined complex sociotechnical systems such as ship navigation (Hutchins 1995a), commercial fishing (Hazlehurst 1994), air traffic control (Halverson 1995), or piloting jet aircraft (Hutchins 1995b; Hutchins and Klausen 1996; Holder 1999), the basic concepts are equally evident in mundane activities like time-telling (Williams 2004). Take, for example the counting of objects as portrayed in the introduction. Counting is a familiar cultural practice for determining quantity. It addresses the question "How many…?" by producing a number that corresponds to a quantity of objects. Determining quantity is a frequently encountered problem for which culture has accumulated partial solutions: various counting practices that use bodily action to coordinate the elements of functional systems. Cultural practices for counting are highly conventionalized, but any situated use of counting must be improvised, in that the form of counting practice, once chosen, must be adapted to and coordinated with the particulars of the setting and situation, including such things as the type and arrangement of the objects to be counted and their physical presence or absence

As an illustration, consider the three common ways to count objects illustrated in Figure 15.1 (analyzed in Williams 2008c). The first, shown in Figure 15.1(a), is to touch objects in succession while uttering number names in memorized sequence: "one, two, three…"; the number name uttered when the last object is touched corresponds to the quantity of objects counted. Here the body provides the coordination necessary for the distributed cognitive system to function: the speech system utters numeric labels in succession while the manual system moves the hand with extended index finger (in the handshape prototypical for pointing) from object to object, touching each object at precisely the instant when a numeric label is uttered. An essential part of this coordination is imposing a path along which the hand moves so that it touches every object exactly once. This system for computing quantity requires the coordination of brain, body, and world (Clark 1997): it combines conceptualization (object perception, a cognitive model for a specific cultural practice, a conceptual path), bodily action (speaking and touching), and environmental structure (a configuration of objects) into a functional ensemble. The system can fail in several ways: missing number names, miscoordinating uttering with touching, mistaking objects to be included, losing track of the counting path, etc. Errors can be reduced through improvised adaptations: a child counting dots in a circle, for example, held the tip of her left index finger at one dot while she used the tip of her right index finger to touch each of the subsequent dots around the circle while counting aloud; marking the start of the counting path in this way made it easier to discern its end (Williams unpublished data). A common adaptation is to modify the environment before counting: to rearrange objects into a line or array in order to facilitate a simpler counting path. These manual actions before counting reduce errors in the execution of the functional system, making it more robust. These examples show how conventional cultural practices must be adapted to the particulars of setting and situation when distributed cognitive functional systems are instantiated in real human activity. Actions of the hands are critical to these situation-specific adaptations.

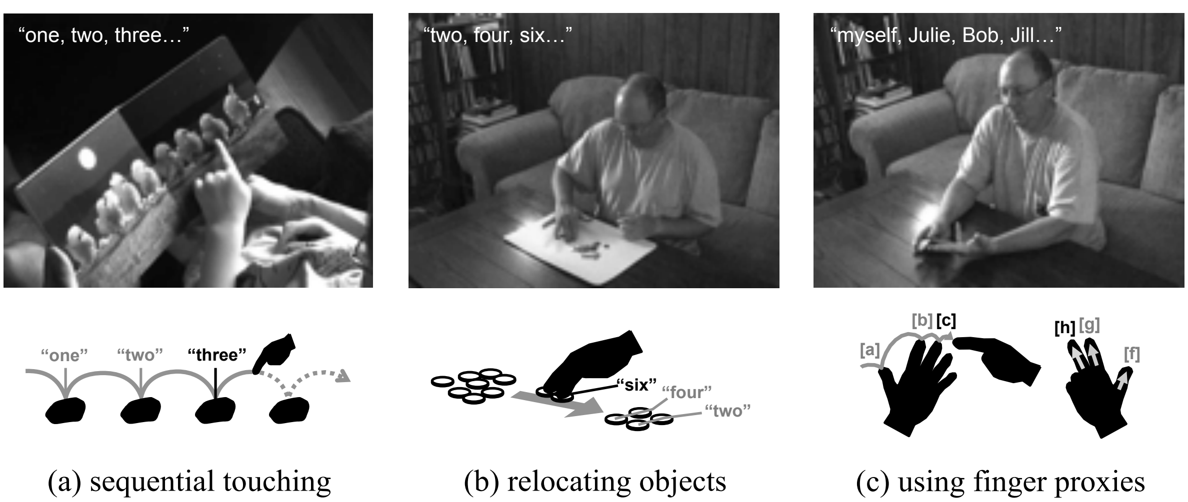

Fig. 15.1: Three functional systems for counting objects

In the form of counting practice described above, the functional system operates through a series of touch-points synchronized with speech; this looks like gestural action without intent to communicate. The gestural quality is even more apparent when the manual actions are reduced to a series of points (with no contact) while number tags are uttered. With further reduction, the system operates with no hand action at all, replacing it with gaze shifting: looking at (fixating) objects in succession while uttering number names subvocally. These are varied instantiations of a common counting practice realized through different bodily actions, some more overtly gestural than others.

A second conventional way to count objects, shown in Figure 15.1(b), involves changing the location of objects while uttering number tags. Examples from Williams (2008c) include picking up and placing traffic cones, sliding coins across a table two at a time, and dropping buttons into a bag. Again coordination of the functional system is achieved through manual action synchronized with speech, but here the movements look more like practical actions than gestures: grasping, lifting, sliding, placing, and dropping objects, all performed not to accomplish some practical end on or with an object but rather to accomplish the cognitive goal of computing quantity.

A third way to count objects, shown in Figure 15.1(c), again appears gestural in form: raising fingers or touching fingers successively (for example, to a surface) while uttering names of non-present entities. Here the fingers are proxies for objects being counted. Examples from Williams (2008c) include raising fingers while reciting the alphabet to identify the 18th letter and touching and raising fingers while naming family members to determine the number of people for a dinner reservation. In this functional system, the hand configuration is modified in coordination with object-naming; the configuration produced when the final object is named represents the total number of objects. This final configuration can be identified using associations from childhood counting practice, or the finger-raising sequence can be repeated while reciting number tags until the target configuration, held in visual working memory, reappears. The manual actions in this case are neither practical nor communicative: they are cognitive actions that encode representational states during the execution of a computational process. If we call them gesture, as I believe we should, then they are gesture for cognition, specifically, gesture for problem-solving rather than word-finding or thinking-for-speaking, which are cognitive functions claimed for gesture when it is considered in relation to speech.

The distributed cognitive functional systems described in this section all accomplish computation through sequenced actions of the hands (or eyes) that coordinate spoken labels with objects or their proxies. The manual actions that bring these systems into coordination cross distinctions between practical action (physically moving objects), communicative action (pointing), and cognitive action (counting on fingers). A single form, an index-finger point, can serve different functions (cognitive, communicative, or both simultaneously) while a single function, coordinating number tags with objects, can be accomplished by different forms (looking, pointing, touching, sliding, picking up and placing, etc.). Whether manipulative or gestural in appearance, the hand actions described here serve the common purpose of coordinating elements in a functional system to produce a computational outcome.

4. The affordances of gesture as a representational medium

Human hands are the first tools of representation. Streeck (2009: 39—58) describes the capabilities of hands that form the basis for practical actions and gestural movements. From the perspective of distributed cognition, hands can represent, can produce representations in other media, and can propagate representational state from one medium to another, including to or from themselves. In many respects, this puts hands at the center of human cognition, with heads as controllers of hands and interpreters of the states they produce. Hands act on and modify the world: they manipulate objects, rearrange them, shape them, assemble or disassemble them, transport them, and employ them as tools to act on other objects (to draw, write, carve, and so on). These are commonly regarded as practical actions, but they may equally be cognitive actions, helping us perceive the affordances of objects (Kirsh and Maglio 1994) or prepare the environment for intelligent action (Kirsh 1995), as in the example of lining up objects before counting them. Hands also interact with the world without modifying it: they bring attention to objects, highlight their relevant features, and annotate or elaborate their structure. These are environmentally coupled gestures (Goodwin 1994, 2007) whose significance derives from the culturally constituted spaces in which they are performed. Finally, hands depict directly, in the space in front of the speaker, using conventional gestural practices: they enact schematic actions; they evoke imagined objects through enactments or through schematic acts of drawing, outlining, or molding; and they model objects and their interactions (see Müller 1998: 114—126 and Streeck 2008 for discussion of gestural modes of depiction). Acting on objects, gesturing over objects, and gesturing in the air seem like different sorts of hand actions, but from the perspective of distributed cognition, they often serve similar or closely related purposes. The purpose of a given hand action may become apparent only when it is considered in terms of the functional system being instantiated to accomplish a particular outcome.

That human hands modify environments, manufacture artifacts, and use tools to achieve desired ends is well known and widely regarded as fundamental to human life. More specific to human cognitive achievements are hands' entrained abilities to create representational states in physical media. Hands create representational states through culturally shared practices of sketching, drawing, and writing, as well as through more specialized practices like painting, sculpting, carving, or crafting. In much of the world today, hands create representational states in electronic media through historically recent practices such as keyboarding, mousing, and using touch-pads and -screens. Where physical or electronic media are absent, or where the skills to employ them are lacking, hands rely upon themselves to represent: they use their own physicality to materialize conceptual content. Indeed, this ability may be integral to all the others. Hutchins (2010) claims that "humans make material patterns into representations by enacting their meanings" (Hutchins 2010: 434), and hands are humans' primary tools for enactment.

Given this array of potential means for representation, it is worth asking: What are the affordances of hands that lead to their being employed for depiction when other media might instead be chosen? Because hands are parts of our bodies, they are always "ready at hand": they can be brought into action quickly and can produce representational states faster than these could be engendered in other media. In contrast with writing and drawing, hands represent relations in three-dimensional space and can move while representing, enacting the dynamics of processes. Multiple changing relations are especially hard to visualize, requiring complicated physical models or clockworks, flat (2-D) video recordings or animations, or high-technology systems for motion capture or 3-D visualization. Hands can conjure 3-D relations and dynamics directly in space—in so far as a partial, schematic depiction annotated by speech is sufficient to the demands of the situation and the complexity of what needs to be represented.

Using the hands depictively also brings processes of all sizes and scopes, from the cosmic to the microscopic, into "human scale": the scale at which we directly perceive and act in the world (Fauconnier and Turner 2002: 312). Two types of human scale are important in gesture research: one in which the gesturer inhabits a space and acts subjectively within it, called "character viewpoint," and another in which the gesturer models objects and interactions in the space in front of his body, called "observer viewpoint" (McNeill 1992: 118—125). A speaker adopts character viewpoint if she enacts steering a car while describing an automobile accident; she adopts observer viewpoint if she uses her hands to model two cars colliding, a depiction she views from outside the space of action in mutual orientation with her interlocutor, who views it from another angle. Observer viewpoint, in particular, enables us to take processes at any imaginable scale and to portray them in the perceivable, reachable space in front of our bodies and thus to "dominate" them (Latour 1986: 21). And because our bodies are mobile — able to bend, reach, turn, walk, and so on — we can transport gestural representations into and out of co-location with states in other media, thereby linking or coupling them. This puts hands as representational tools squarely at the center of the coordinative processes necessary for cognitive functional systems to achieve their outcomes.

Finally, what must be noted in this discussion of the affordances of gesture as a representational medium is the limited durability, the non-persistence, of gestural representations. Gestures have a greater material presence than words, but while they can be sustained briefly to support perception and reasoning, they vanish as soon as the hands are put to other uses. This is their most significant contrast with other physical media, yet the affordances of gesture enable it to be used in conjunction with durable media to achieve outcomes more powerful than either could achieve on its own.

5. Using hands to create and coordinate representational states

The examples discussed below are taken from studies of distributed cognition in various settings. They demonstrate how gesture is used to represent and to coordinate representations in functional systems for accomplishing cognitive activity.

5.1 Creating provisional representations during joint imagining

An example that illustrates the coordination of gesture with other representational media comes from the situation of naval quartermasters plotting lines of position on a navigation chart (Hutchins 1995a, 2010). A navigation chart is a computational device: a line drawn on the chart restricts the possible locations of the ship with respect to the surroundings; a second, intersecting line determines this location uniquely; in standard practice, a third line forms a small triangle whose magnitude corresponds to the margin of error or indeterminacy in the position fix (Hutchins 1995a: 61). The navigation chart itself incorporates the residua of cognitive processes extending across multiple scales of time, from the momentary actions of the navigation team plotting the fix; to the earlier actions of team members updating the chart; to the actions of others, distributed across time and space, mapping the world represented in the chart; to the origins of the representational systems and conventions (latitude/longitude, compass directions, and so on) that enable the outcomes of mapping expeditions to be combined; to basic practices of counting and measurement whose roots lie in the ancient world. These different timescales of activity are apparent in the means through which representational structure is layered on the chart (Hutchins 1995a: 165—167). The outcomes of centuries of past activity are captured in the printed features of the chart: in its lines, shapes, scales, and labels. Updates to the chart (new landmarks or underwater hazards, turn bearings or danger depth contours, etc.) are added in ink before the chart is employed in navigation. When the chart is in use, plotted lines of position and projected future positions are marked in pencil. Finally, when navigators consider possible landmarks for the next position fix, they trace projected lines of position on the chart with their index fingers, leaving no marks. The significance of these gestural traces emerges not simply from the composites of gesture and speech (the utterances) but from the layering of the gestures, construed by speech, on the meaningful space of the chart in the context of the mutually understood activity being jointly pursued. As Hutchins (2006) notes: "The meanings of elements of multimodal interactions are not properties of the elements themselves, but are emergent properties of the system of relations among the elements" (Hutchins 2006: 381). These gestural traces of possible lines of position are part of an embodied process of joint imagining: perceiving in a "hypothetical mode" (Murphy 2004: 269) while acting in a "subjunctive mood" (Hutchins 2010: 438). The fleeting quality and lack of physical imprint of the gestures suit precisely the nature of the cognitive task at hand: considering, but not committing to, possible courses of action, and using these considerations as the basis for a decision that will determine future action.

5.2 Adding a third dimension and motion dynamics for scientific visualization

In the example of tracing imaginary lines of position on a navigation chart, gesture layers structure onto an existing material representation, adding constraint to isolate an outcome. In the next example, from a meeting in a scientific laboratory (Becvar, Hutchins and Hollan 2005), gesture extracts representational state from a flat representation and transforms it, adding a third spatial dimension and motion dynamics. The result is a human-scale, hand-based model for theorizing about molecular interactions, again in a hypothetical mode. In this case, the principal investigator in a chemistry laboratory has just projected a ribbon diagram of the thrombin molecule on an overhead projector, as shown in Figure 15.2(a). She calls attention to the many loops in the diagram ("see how you have all these little loops: this loop, this loop, this loop, and this loop"), pointing to examples on the transparency with her left index finger, silhouetted on the projected image, as she identifies them. She begins to say "all kin' of" and then breaks off her speech, whereupon she lifts her left hand from the transparency into the air, palm outward with fingers outstretched, and says "in three-dimensional space they're like this"; this moment is shown in Figure 15.2(b). Her left hand has extracted representational state, the basic morphology of the thrombin molecule, from the ribbon diagram and reproduced it in three dimensions in the space in front of her body. Her hand *is* the molecule, and her fingers now represent the loops she has identified. She holds this hand configuration just below eye level in clear view of the audience she is addressing, a typical position in gesture space for depictive (iconic) gestures intended for mutual orientation. Then she elaborates this 3-D model. First, she points with her right index finger to a location between the finger-loops, saying "an' that's the active site"; here her left hand models the molecule using a body-part-as-object form of depiction, while her right hand indexes a specific site on this molecule using the conventional form of index-finger pointing associated with individuating a reference object or location. This two-handed gesture complex, in coordination with the accompanying speech, accomplishes the multifaceted purpose of representing molecular structure, making it available for visual scrutiny, and focusing attention on a detail in that structure that is critical to understanding how the molecule functions. In a further elaboration, the speaker adds motion dynamics to the 3-D model. She says, "And so our new theory is that thrombomodulin does something like this," pausing briefly to contract and expand her fingers, "or like this," pausing again to rotate her fingers from side to side. In this portion of the discourse, the speaker uses her hand-as-molecule to enact and thereby simulate possible forms of molecular motion resulting from the binding of thrombomodulin. These simulations are, like the lines-of-position example, not yet committed to but hypothetical. In subsequent elaborations, she places the back of her right hand against the back of the hand-as-molecule to enact the binding of thrombomodulin to the back side of thrombin, and then she uses rapid movement of her right hand toward the interior of the hand-as-molecule to enact the rapid binding of another protein to the active site. Throughout these depictions, her left hand models the thrombin molecule and its dynamics while her right hand alternates between indexing parts of the molecule and modeling other molecules' interactions with it.

(a) thrombin diagram (b) thrombin hand model (c) thrombin hand model (6 months later)

Fig. 15.2: Representational gesture as a cognitive artifact for scientific reasoning

This example clearly illustrates how gestural depiction becomes a component of scientific reasoning. The speaker's hand-as-molecule gesture creates a stable, visually accessible, dynamically reconfigurable, 3-D model of a functioning molecule at human scale. Her gestural elaborations of that model serve to highlight invisible elements and depict imperceptible processes, all in the hypothetical mode. By using her hand movements to create these representations, the speaker also brings her own embodied experience with tangible objects, felt movements, and visuospatial perception into play, providing a bodily basis for sensing connections or making discoveries about molecular dynamics. The gestural model takes on the status of a cognitive artifact: a representational or computational tool that is part of a cognitive functional system—in this case, a system for reasoning about molecular interactions. This gestural model proves crucial to the work of the group, as evidenced by two observations: first, that other members of the group reproduce the hand-as-molecule gesture during subsequent discussion, and second, that the hand-as-molecule gesture is produced independently and spontaneously by a lab member (not the original speaker) six months later in an interview when she describes the goals of the project, as shown in Figure 15.2(c). In both the science laboratory and ship navigation examples, gestural enactments serve as components of hypothetical thinking as well as ways of sharing that thinking with others, demonstrating how hands are used as tools for reasoning as well as communication.

5.3 Coordinating representational states to construct a shared object of knowledge

Another study of a science lab (Alač and Hutchins 2004) reveals additional ways in which gestural actions participate in processes of thinking as well as communicating. Here the focus is on functional magnetic resonance imaging (fMRI) researchers interpreting brain images. As Goodwin observes in "Professional Vision" (1994): "An event being seen, a relevant *object of knowledge*, emerges through the interplay between a *domain of scrutiny*… and a set of *discursive practices*… being deployed within a *specific activity*" (Goodwin 1994: 606: emphasis in original). In the functional magnetic resonance imaging lab, the domain of scrutiny consists of color images displayed on computer monitors; these images have been produced by scans of participants' brains as they viewed visual stimuli. The researchers employ a set of discursive practices, including gestures, to transform these images into objects of knowledge, namely, into representations of brain areas and their levels of activity. As is often the case in these kinds of studies, the process is laid bare through the interaction between an expert and a novice, where the expert teaches the novice how to "see" the phenomena of interest—that is, how to enact a functional system through which the objects of knowledge are made manifest. Processes that for experts are largely covert (though not entirely internal) are made overt in these interactions, and processes of thinking are opened up into processes of communicating, with frequent production of gesture.

In contrast with the preceding example, gesture here is employed less as a means for directly representing than as a means for coordinating representational states in different media: in the brain images displayed on the computer screen; on a paper chart with hand-drawn diagrams of the visual field (what the participant saw) and of retinotopy space (a map of visual cortex); and in the talk produced by the expert as she draws, gestures, and engages with her interlocutor. In one such coordinative sequence, the expert: (1) touches the brain image on the computer screen while identifying the location she touches as the "center"; (2) rotates the hand-drawn chart to align it with the image and points to a location on the chart she identifies as "right here"; (3) holds the chart up next to the computer screen while saying "and when we look at this map it looks something like that"; (4) traces the outline of the primary visual area on the chart with her index and middle finger while saying "so V1 is going to be in the center"; and then (4) transposes her hand — maintaining the tracing handshape — to the brain image on the computer screen where she executes a matching two-fingered trace, shown in Figure 15.3(a), six times in rapid succession while saying "it's gonna be this pie shape; it's probably covering approximately this area" (Alač and Hutchins 2004: 646). Here the coordinative function of gesture is quite apparent. Pointing and tracing highlight structures in external representational media whose conceptual identity is invoked by speech. Maintaining handshape while moving the hand from one culturally constructed space (the chart) to another (the computer image) and repeating the gesture form in the new space together establish a conceptual link between the highlighted states of the two media; these states are construed, named, and related by the accompanying speech ("this," "here," "the center," "V1," "like that," "so," etc.) to produce the relevant object of knowledge, namely, seeing part of the colored image *as* V1, the primary visual cortex. "Seeing-as" is a cognitive accomplishment, the outcome of a discursive process in which gesture weaves conceptual content into material patterns in the physical world.

(a) outlining gesture transposed from chart to image

(b) squeezing gesture overlaid on chart (palms moving inward)

Fig. 15.3: Gestures propagating and coordinating representational states

In other parts of the interaction, gesture is used depictively, as in the hand-as-molecule example, to add dimensionality and dynamics to what is being represented. In Figure 15.3(b), the expert places her hands, wrists together, palms inward, fingers outstretched in a V-shape, on top of bold lines on the chart of retinotopy space as she says "take these two meridians," and then she reduces the angle between her palms as she says "as if you were squeezing them together into the pie shape" (Alač and Hutchins 2004: 642), where "the pie shape" refers to a wedge outlined on the chart. Here the conjunction of chart, hands, and speech construes the action as simulating the inward movement of the meridian lines as they are squeezed to the boundaries of a section in retinotopy space. As Alač and Hutchins point out, this squeezing has "no real-world referent" in that it corresponds to "no real action wherein the visual space is effectively squeezed and transformed into the retinotopically organized visual areas" (Alač and Hutchins 2004: 643). The enacted squeezing is a metaphorical action: a human-scale physical enactment of a more diffuse, invisible process through which the visual field comes to be represented in visual cortex. The novice is directed to imagine the action "as if" it were happening while the gesture enacts the process, as in the earlier examples, in a hypothetical mode.

In the examples described above, meanings are produced not additively, as the aggregation of meanings in separate media, but emergently, as the interrelation of states in media brought into coordination through gesture. Gesture coordinates with visible media through co-location and with audible media through co-timing, while it can also coordinate with both through perceived similarities of form. Proximity and synchrony are precise achievements of skilled human action that weaves media into meaning. As hands enter culturally constituted spaces, they form shapes and perform movements that take on meaning in relation to the structures and representational conventions that govern those spaces. How the gestures are to be interpreted, whether in relation to an artifact itself or to what that artifact represents, depends upon the construal provided by speech and by shared knowledge of the situation. In Hutchins and Palen (1997), for example, a pilot's gestures on and over an instrument panel in the flight deck are variously interpreted as actions taken on the panel and as events occurring in the aircraft's fuel system (Hutchins and Palen 1997: 37). Hands also transport meaningful state from one constructed space — one semiotic field — to another, as when the functional magnetic resonance imaging researcher transfers a traced outline from the labeled chart to the brain image. Moving a representational state into a new semiotic field transforms how that state is seen. Hutchins gives the example of a navigator moving his dividers (a tool that can be set to span a particular interval) from a line segment on a navigation chart to a printed scale where the distance traveled in three minutes (1500 yards) can be read as a speed (15 knots) by ignoring the two trailing zeroes (1995a: 151—152; further analyzed in 2010: 429—434). Whether what is moved from one space to another is a physical tool or a configured hand makes little difference; what matters is the semiotic shift. Of course, hands also create representations in their own semiotic space, in the air in front of the speaker, in relation to spoken content, as when the scientist uses her hand to model the shape and movements of a molecule. Through these processes, hands play a crucial role in producing and elaborating "multimodal meaning complexes" (Alač and Hutchins 2004: 637) in the interactions through which joint activities are accomplished and through which participants come to share understanding.

5.4 Coordinating representational states across multiple participants

The multimodal meaning complexes described up to this point appear to result from the actions of a single participant at a time, but this need not be the case. In the examples of presenting to a group and of teaching a novice, the speaker is more knowledgeable, is acting in the expert role, and is holding the floor during the analyzed segment of discourse. In the work situations studied by Hutchins, Goodwin, Streeck, and others, participants frequently engage in familiar activities in known settings with mutually understood goals and overlapping knowledge. The resulting high degree of intersubjectivity and more balanced participation make it increasingly likely that the gestures and speech of different participants will mutually elaborate one another. In an examination of three pilots interacting in a training situation where one (American) is instructing the other two (Japanese) in a Boeing aircraft procedure, Hutchins and Nomura (2011) find multiple instances where gesture produced by one participant develops meaning in relation to talk produced by another. In the title of their paper, Hutchins and Nomura refer to this phenomenon as "collaborative construction of multimodal utterances." In the cases they describe, the collaborative construction is directed toward creating a shared conceptual object: a sequence, cause/effect, comparison, etc. The gestures produced by one pilot while another speaks simulate interaction with aircraft control systems by enacting virtual actions on imagined objects, or else they model aircraft responses using the hands or body with outstretched arms to simulate changes in the airplane's orientation or dynamics. The gestural enactments variably precede (with a hold), coincide with, or quickly follow their lexical affiliates in the other's speech. Their purpose seems to be to display intersubjective understanding through demonstrated action (in anticipation of or in response to the other's verbalization) and/or to practice the procedure being described, possibly as an aid to future recall. Listener gestures provide visible evidence that the listener inhabits a conceptual world in common with the speaker. In so doing, they require a commitment to particulars of the situation not evident in speech. Gestural enactments evoke an imagined setting (the flight deck of a jet aircraft), a role (pilot flying), and a vantage point (in the right or left seat), including details such as the location and operation of aircraft controls. Aspects of the setting "are brought forth as implied elements in an imagined world of culturally meaningful action" (Hutchins and Nomura 2011: 41), and gestures "are coupled to elements of that imagined environment" (Hutchins and Nomura 2011: 42). The speaker also appears to modify his on-going talk as a consequence of the other's gestures, variously confirming or correcting the apparent interpretations or omitting items that have already been established, such that a lexical affiliate for the other's gesture may never be spoken. Hutchins and Nomura claim that "the participants are engaged simultaneously in two kinds of projects: they are enacting conceptual objects of interest (what they are talking about), and they are conducting a social interaction. While these objects are analytically separable, in action, they are woven into the same fabric" (Hutchins and Nomura 2011: 40). In these examples, talking and gesturing in coordination with others appears to be a way of establishing common ground for the discourse while simultaneously establishing the objects of knowledge that define the work of the profession.

Finally, it is worth pointing out that the most significant phenomena described here — the coupling of gesture with other media, and the coordinated production of talk and gestures by separate participants — are precluded by an experimental methodology in which one participant who has seen a video narrates the events, without access to any material resources, to another who has not. These phenomena are also less likely to be observed in studies of conversations where participants tell stories about people, happenings, and objects not present. This point is not meant to diminish the many insights that are gained from such studies. It is, however, meant to press the claim that the primordial home for gesture is in mutual, consequential activity in culturally constituted settings. Gesture in such activity is likely to be performed in relation to other media and in close interaction with other participants, and it is likely to serve a functional role in cognition that goes beyond the expression of internal content.

6. Using hands to propagate functional systems across generations

We have considered ways in which people use their hands to create and coordinate representational states to accomplish cognitive activities, individually and in collaboration with others. With frequently recurring tasks, these practices can become highly conventionalized, although, as we have seen, they are always adapted to the particulars of situations. Taking a broader perspective, we may now ask: How are distributed cognitive functional systems — coordinations of cognitive models, artifacts, and cultural practices — propagated across generations? Here again gesture plays a crucial role.

We have already seen one illustration of the propagation of cultural practices in Alač and Hutchins' (2004) example of the experienced functional magnetic resonance imaging researcher teaching the novice how to interpret brain images. Here the expert used pointing and tracing to highlight shapes in the visible media (the chart and the image) while her speech profiled conceptual entities and relations manifested in those shapes. Keeping a fixed handshape while moving her hand from one semiotic space (the chart) to another (the image) and repeating the gestural form helped establish a conceptual link between elements in the two spaces. By compressing analogy into identity (Fauconnier and Turner 2002: 314-–315, 326), the novice learns to see the shape on the computer screen as a cortical map in the participant's brain. It may be that the expert routinely uses her hands to accomplish this seeing on her own, perhaps by tracing outlines on the images she is examining, but when she teaches the novice, she performs these actions overtly, opening the process to scrutiny. She also annotates her task-relevant actions with additional gestures, speech, and shifts of gaze from work objects to her addressee, monitoring the novice's responses as she explicitly guides him in where to look (how to attend), what to see (how to conceptualize what is being viewed), and what to do (how to act). The expert shifts projects from accomplishing to teaching. This shift is evident in how she orients her body and how she uses her hands and her talk to engage her interlocutor as well as the tools of her trade. In her instructional discourse, the expert demonstrates how to find, interpret, coordinate, and employ relevant states of representational media to accomplish the work of a functional magnetic resonance imaging researcher.

A more commonplace example of shifting projects from doing to teaching, of opening functional systems to scrutiny, and of guiding the conceptualization of novices can be found in adults teaching children to tell time (Williams 2008a, b). Expert time-tellers look at an analog clock and read the time with gaze-fixing and slight gaze-shifting from one clock hand to another as the only visible evidence of a cognitive process unfolding. It is doubtful that any novice could learn to read the clock simply by watching an expert do it. Children learn to read the clock because adults who are proficient time-tellers provide them with active instruction: pointing to structures and tracing paths on the clock face, highlighting elements, relations, and processes while construing them with speech, and shifting gaze from the clock face to the learner to monitor attention and seek signs of confusion or comprehension. While the child practices reading times on the clock, the adult monitors and provides prompts, confirmation or correction, and additional instruction as needed. Through this form of social interaction, children learn to see meaningful structure on the clock face and to interpret that structure in relation to human activity and to a conventional system of time measurement. Seeing time on the clock is another cognitive accomplishment entrained by the gestural weaving of material and conceptual worlds.

As a brief example, consider the fragment of instruction shown in Figure 15.4 (Williams 2008a). Here the teacher says "now another way that we say it, is we count by fives, when we move this from number to number; there's five minutes between each number" while she enacts a hypothetical process of counting on the clock. If we break this fragment into segments, we see the dynamic mapping of conceptual content to the clock face as mediated by gesture. While saying "now another way that we say it," the teacher moves the minute hand to the 12, positioning the hand at the starting point for a clock-counting process. When she says "is we count by fives," she activates a cognitive model for counting that is familiar to her first grade class: touching objects (sets of five elements) while uttering "five, ten, fifteen…."

Accompanying her statement "when we move this" is a shift of gaze to where the tip of her right index finger rests on the minute hand; this construes the minute hand as the thing-to-be-moved, namely as the pointing finger that will touch each object-set as it is counted. The next part of her utterance, "from number to number," defines numbers as the object-sets to be counted; here she touches the large numerals on the clock face in sequence, making it clear which numbers she is referring to, while the form of her gestural movement enacts a canonical counting motion, bouncing from one number to the next clockwise around the dial. The gesture alone provides the origin, direction/path, and manner of the counting motion, which is notably not the continuous, steady movement of a clock hand but the intermittent, bouncing movement of a human hand touching objects while counting. The same gesture continues during the next statement, "there's five minutes between each number," a statement that activates a cognitive model for the conventional system of time measurement, in which an hour is divided into 60 minutes, and maps an interval of five minutes to the space between adjacent numbers on the clock. In this example, a single gesture in coordination with two verbal statements sets up mappings from two cognitive models: a mapping from objects in the counting model to numbers on the clock face, and a mapping from units of time in the time measurement model to intervals of space on the clock face. The second mapping conjoins with the first to implicitly generate a third mapping: linking units in the system of time measurement (minutes) to elements of the object-sets being counted (five minutes per object-set). All of this is accomplished through the coordination of gaze, gesture, speech, and a culturally constituted artifact, all carefully orchestrated to guide the novice's conceptualization (Williams 2008b). Once these mappings are established, the teacher performs the counting process by grasping the minute hand and moving it to the 5, the 10, and the 15, pausing momentarily at each while saying "five, ten, fifteen…." If the children have succeeded in making the correct conceptual mappings, they will see the clock hand as a counting finger that touches each number in sequence while the elapsed minutes are counted. This, in microcosm, is how conventional functional systems get propagated, sustaining the cognitive accomplishments of the human species.

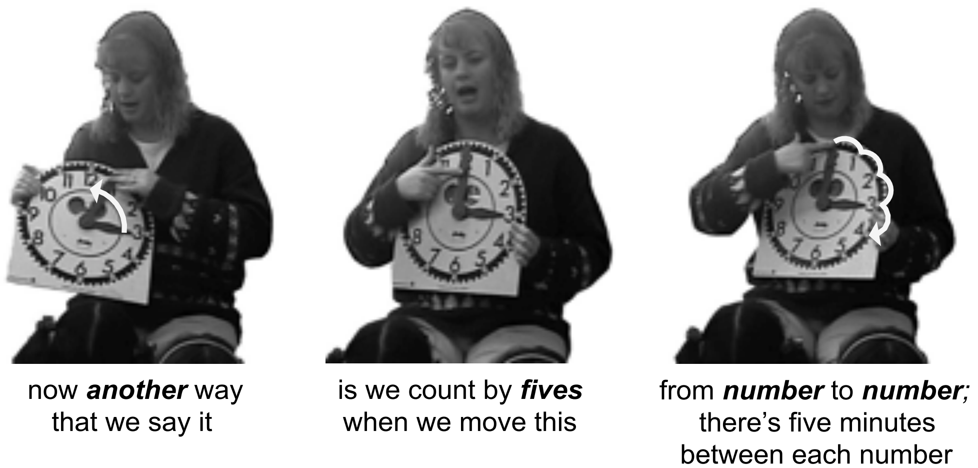

Fig. 15.4: Gestures guiding conceptual mapping

Given this instruction, the children must perform the activity with diminishing help until they are able to instantiate the functional system successfully in appropriate contexts with little effort; only then would we say that they have mastered the practice. Once they become proficient and use the system repeatedly, they will come to recognize the hand configurations and numeric labels as standing for particular five-minute times (oh five, ten, fifteen, and so on), and they will shift strategies from counting to directly naming these times, retaining counting as a backup strategy should memory fail them. A new functional system will emerge, one that supports more efficient conduct of the activity while it reduces the cognitive demands on the individual coordinating the system to produce the intended outcome. The expert system will differ from the novice system, but the counting-based practice will continue to be retained by our culture as a stepping-stone because it enables the sustained successful performance through which the memory-based ability arises.

7. Conclusion

This article has presented evidence for gesture's role in: (1) coordinating the functional systems through which cognitive work gets done, and (2) propagating those systems across generations. In purposeful human activity, participants gesture not simply to express but to accomplish. The familiar conduit metaphor of communication proves inadequate for studying meaning-making in situated activity because it obscures the ways gesture operates in distributed systems for human cognition. Even where the focus of study is exclusively on speech, the conduit metaphor tends to mislead because, as Hutchins (2006) points out, “it is easier to establish a meaning for words embedded with gestures that are performed in coordination with a meaningful shared world than it is to establish meanings for words as isolated symbols” (395). That humans can communicate solely through words is clear, but that such communication should be regarded as prototypical is clearly mistaken. Recognizing this, leading gesture researchers like Kendon and McNeill have argued that gesture, like speech, is part of utterance. Researchers who study distributed cognition find it more productive to treat gesture as part of the functional systems through which cognitive outcomes are accomplished. If we expand the unit of analysis to encompass aspects of the setting, of mutual orientation and (inter-)action, and of shared knowledge and the unfolding of goal-directed activity, then we stand a better chance of understanding and appreciating the critical role that gesture plays in human cognition and communication.

8. References

Alač, Morana and Edwin Hutchins 2004 I see what you are saying: Action as cognition in fMRI brain mapping practice. *Journal of Cognition and Culture* 4(3): 629—661.

Becvar, L. Amaya, James Hollan and Edwin Hutchins 2005 Representational gestures as cognitive artifacts for developing theory in a scientific laboratory. *Semiotica* 156(1/3): 89—112.

Chafe, Wallace 1994 *Discourse, Consciousness, and Time: The Flow and Displacement of Conscious Experience in Speaking and Writing*. Chicago: University of Chicago Press.

Clark, Andy 1997 *Being There: Putting Brain, Body, and World Together Again*. Cambridge, MA: Massachusetts Institute of Technology Press.

Fauconnier, Gilles and Mark Turner 2002 *The Way We Think: Conceptual Blending and the Mind's Hidden Complexities*. New York: Basic Books.

Goodwin, Charles 1994 Professional vision. *American Anthropologist* 96(3): 606—633.

Goodwin, Charles 2000 Action and embodiment within situated human interaction. *Journal of Pragmatics* 32: 1489—1522.

Goodwin, Charles 2007 Environmentally coupled gestures. In: Susan D. Duncan, Justine Cassell, and Elena T. Levy (eds.), *Gesture and the Dynamic Dimension of Language: Essays in Honor of David McNeill*, 195—212. Amsterdam/Philadelphia: John Benjamins.

Halverson, Christine 1995 Inside the cognitive workplace: New technology and air traffic control. Ph.D. dissertation, Department of Cognitive Science, University of California, San Diego.

Hazlehurst, Brian 1994 Fishing for cognition: An ethnography of fishing practice in a community on the west coast of Sweden. Ph.D. dissertation, Departments of Anthropology and Cognitive Science, University of California, San Diego.

Holder, Barbara 1999 Cognition in flight: Understanding cockpits as cognitive systems. Ph.D. dissertation, Department of Cognitive Science, University of California, San Diego.

Hutchins, Edwin 1995a *Cognition in the Wild*. Cambridge, MA: Massachusetts Institute of Technology Press.

Hutchins, Edwin 1995b How a cockpit remembers its speeds. *Cognitive Science* 19: 265—288.

Hutchins, Edwin 2001 Distributed cognition. In: Neil J. Smelser and Paul B. Baltes (eds.), *International Encyclopedia of the Social & Behavioral Sciences*, 2068—2072. Oxford: Elsevier.

Hutchins, Edwin 2003 Cognitive ethnography. Plenary address at the 25th meeting of the Cognitive Science Society, Boston, MA, July 31–August 2.

Hutchins, Edwin 2006 The distributed cognition perspective on human interaction. In: Nick J. Enfield and Stephen C. Levinson (eds.), *Roots of Human Sociality: Culture, Cognition and Interaction*, 375—398. Oxford/New York: Berg.

Hutchins, Edwin 2010 Enaction, imagination, and insight. In: John Stewart, Olivier Gapenne, and Ezequiel A. Di Paolo (eds.), *Enaction: Towards a New Paradigm for Cognitive Science*, 425—450. Cambridge MA: Massachusetts Institute of Technology Press.

Hutchins, Edwin and Tove Klausen 1996 Distributed cognition in an airline cockpit. In: Yrjö Engeström and David Middleton (eds.), *Cognition and Communication at Work*, 15—34. New York: Cambridge University Press.

Hutchins, Edwin and Saeko Nomura (2011) Collaborative construction of multimodal utterances. In: Jürgen Streeck, Charles Goodwin, and Curtis LeBaron (eds.), *Multimodality and Human Activity: Research on Human Behavior, Action, and Communication,* 29—43. Cambridge: Cambridge University Press-

Hutchins, Edwin and Leysia Palen 1997 Constructing meaning from space, gesture, and speech. In: Lauren B. Resnick, Roger Säljö, Clotilde Pontecorvo, and Barbara Burge (eds.), *Discourse, Tools, and Reasoning: Essays on Situated Cognition*, 23—40. Berlin: Springer-Verlag

Kendon, Adam 1972 Some relationships between body motion and speech: An analysis of an example. In: Aaron Siegman and Benjamin Pope (eds.), *Studies in Dyadic Communication*, 177—210. Elmsford, NY: Pergamon Press.

Kendon, Adam 1980 Gesticulation and speech: Two aspects of the process of utterance. In: Mary Ritchie Key (ed.), *The Relationship of Verbal and Nonverbal Communication*, 207—227. The Hague: Mouton and Co.

Kendon, Adam 2004 *Gesture: Visible Action as Utterance*. Cambridge/New York: Cambridge University Press.

Kirsh, David 1995 The intelligent use of space. *Artificial Intelligence* 73: 31—68.

Kirsh, David and Paul Maglio 1994 On distinguishing epistemic from pragmatic actions. *Cognitive Science* 18(4): 513—549.

Latour, Bruno 1986 Visualization and cognition: Thinking with eyes and hands. *Knowledge and Society: Studies in the Sociology of Culture Past and Present* 6: 1—40.

LeBaron, Curtis D. and Jürgen Streeck 2000 Gestures, knowledge, and the world. In: David McNeill (ed.), *Language and Gesture*, 118—138. Cambridge/New York: Cambridge University Press.

McNeill David 1985 So you think gestures are nonverbal? *Psychological Review* 92(3): 350—371.

McNeill, David 1992 *Hand and Mind: What Gestures Reveal About Thought*. Chicago: University of Chicago Press.

McNeill, David 2005 *Gesture and Thought*. Chicago: University of Chicago Press.

Müller, Cornelia 1998 *Redebegleitende Gesten: Kulturgeschichte – Theorie – Sprachvergleich*. Berlin: Arno Spitz.

Murphy, Keith M. 2004 Imagination as joint activity: The case of architectural interaction. *Mind, Culture, and Activity* 11(4): 267—278.

Reddy, Michael J. 1979 The conduit metaphor: A case of frame conflict in our language about language. In: Andrew Ortony (ed.), *Metaphor and Thought*, 284—297. Cambridge: Cambridge University Press.

Streeck, Jürgen 2008 Depicting by gesture. *Gesture* 8(3): 285—301.

Streeck, Jürgen 2009 *Gesturecraft: The Manu-facture of Meaning*. (Gesture Studies 2.) Amsterdam/Philadelphia: John Benjamins.

Williams, Robert F. 2004 Making meaning from a clock: Material artifacts and conceptual blending in time-telling instruction. Ph.D. dissertation, Department of Cognitive Science, University of California, San Diego.

Williams, Robert F. 2006 Using cognitive ethnography to study instruction. In: Sasha A. Barab, Kenneth E. Hay, and Daniel T. Hickey (eds.), *Proceedings of the 7th International Conference of the Learning Sciences, Vol. 2*, 838—844. International Society of the Learning Sciences (distributed by Lawrence Erlbaum Associates).

Williams, Robert F. 2008a Gesture as a conceptual mapping tool. In: Alan Cienki and Cornelia Müller (eds.), *Metaphor and Gesture*, 55—92. (Gesture Studies 3.) Amsterdam/Philadelphia: John Benjamins.

Williams, Robert F. 2008b Guided conceptualization: Mental spaces in instructional discourse. In: Todd Oakley and Anders Hougaard (eds.), *Mental Spaces in Discourse and Interaction*, 209—234. Amsterdam/Philadelphia: John Benjamins.

Williams, Robert F. 2008c Situating cognition through conceptual integration. Paper presented at the 9th conference on Conceptual Structure, Discourse, and Language, Case Western Reserve University, Cleveland, OH, October 18—20.

Robert F. Williams, Appleton, WI (USA)