8

Gesture, Language, and Cognition

Kensy Cooperrider and Susan Goldin-Meadow

8.1 Introduction

Sounds in the air, words in a transcript, rules in the head. These are the phenomena that researchers in linguistics have always focused on. As a result, we now know a great deal about the remarkable formal properties of human language. But this is only one side of language, and another side is easy enough to see. Next time you find yourself in a lively public place, tune out what the people around you are saying and instead watch what they are doing with their bodies. Where you are does not really matter: it could be a quad in California, a coffee shop in Istanbul, a village in Papua New Guinea; the people could be speaking Japanese, Quechua, Russian, or any other language. If you cannot make out what the people are saying or do not understand the local language, all the better. With the words tuned out, another side of language will come to life. Speakers will point, shrug, dismiss, affirm, and implore; they will use their hands to re-enact events, or to depict the shape and size of invisible objects; they will slice, punch, poke, and pull at empty air. Granted, it may not be obvious what these conversations are about. Perhaps a complicated moral quandary? A remodeling project? A close call? But the take-away should be clear: Whenever and wherever people talk, no matter the language or the topic, they gesture. aLanguage and gesture go hand in hand from a very young age, and many speculate that they have gone hand in hand since the earliest stirrings of communication in our species. How does gesture's pervasiveness change our understanding of language and its cognitive underpinnings?

One of the guiding ideas of cognitive linguistics is that language is bound up with cognitive systems for acting, perceiving, attending, reasoning, and so on. Early evidence for this idea was found in the structures of language, how those structures are put to use, and how they change over time (Lakoff and Johnson 1980a, Langacker 1987, Sweetser 1990, Talmy 2000a). More

recent evidence has come from psycholinguistic experiments that shed light on what is going on in the mind when people produce or understand language (Bergen 2012). Our focus in this chapter is on another source of evidence for the interplay of language and cognition: the gestures – like those described above – that universally, ubiquitously, and spontaneously accompany speech. When we look closely at how gesture partners with words, it becomes clear that it is much more than some bodily version or 'translation equivalent' of spoken language. Gesture is a distinctive medium of expression. In what follows, we show that examining gesture can provide insights into language and cognition. In section 8.3, we focus on gesture's intimate relations with spoken language, as well as gesture's ability to reveal aspects of conceptualization that are not evident from the words in a transcript. In section 8.4, we describe the communicative functions of gestures – that is, how gestures serve to convey important layers of meaning to listeners. In section 8.5, we describe the cognitive functions of gestures – that is, how gesturing helps speakers talk and think, and how it shapes reasoning. Before we can consider gesture's relations to language and cognition, however, we must first consider some preliminaries: what counts as a gesture, how we know when a gesture starts and stops, and how gestures can be classified once they have been identified.

8.2 Defining, Individuating, and Classifying Gestures

In popular parlance, the term *gesture* is used loosely and unevenly. In the academic study of gesture, however, the term has a restricted meaning: Gesture refers to movements of the body that are produced in the course of communication, as part of utterances (often called 'co-speech gestures') or even as utterances in themselves (Kendon 2004). This definition excludes practical, non-communicative movements, like taking a sip of coffee or fixing a ponytail. These are not integrated with speech in the same way and lack other properties of gesture. Research on gesture has focused on gestures of the hands, though people do gesture with other body parts, especially the head and face. Headshakes and head nods are commonplace if not universal (Jakobson 1972, Kendon 2002) and, in some cultures, the head and face play a prominent role in pointing (e.g. Sherzer 1972). When we say gesture in what follows, we mean gestures of the hands unless otherwise specified.

Gesture can sometimes look like a blur of movement. But despite appearances, the gestural stream can be segmented into individual units and subparts, much like the speech stream. Here we cover only the very basics of gesture segmentation (more detailed discussions can be found in McNeill 1992, and Kita, Van Gijn, and Van der Hulst 1998). The heart of a gesture is called the *stroke* – this is the part of the gesture that is most

meaningful and effortful. Other movements play important supporting roles too. In order to produce a gesture stroke, a speaker will often first have to get the hand or hands into better position. This phase is called the *preparation*. For some gestures, once the stroke has been produced, the speaker's hand freezes in place. This phase is called a *hold*. Finally, after a gesture stroke is completed, the hands return to a resting position. This phase is called the *retraction*. In an unbroken sequence of gesturing, the stroke of one gesture will give way to the preparation phase of the next, with no retraction until the end of the entire sequence. Note that the stroke of the gesture is really what we focus on when we classify gestures, and when we consider the relationship between a gesture and its accompanying speech.

A number of different schemes have been used to classify gestures into different types (a history of these schemes is reviewed in Kendon 2004). Many of the schemes are grounded in a basic distinction that the philosopher Charles S. Peirce made between the three different ways that a sign can carry meaning. Here a *sign* is any communicative unit, whether a gesture, a word, a graphical symbol, or the like (for discussion, see Clark 1996, Enfield 2009). Peirce's categories are useful not only for classifying gesture, but also for thinking about how gestures relate to spoken and signed language, as well as to other communicative forms.

First, gestures can mean by creating a direct connection to a referent in space and time – this meaning relation is sometimes called *indexicality* or simply *indicating*. The canonical indicating gesture involves extending the arm and index finger, but people also point with a jerk of the head or in variety of other ways across cultures (see contributions in Kita 2003). Tapping or holding up an object that one is talking about are also ways of indicating (Clark 1996).

Second, gestures can mean by looking like what they refer to – that is, by depicting a referent in at least some respects. This meaning relation is sometimes called *iconicity*, and gestures excel at it. Because gestures involve actions of the body, they readily depict human actions. Similarly, because they unfold in space, they excel at depicting motion through space and spatial relationships (Emmorey, Tversky, and Taylor 2000). Gestures are not limited to representing actual actions, concrete images, and real spatial relationships, however. They are equally good at embodying metaphors in which actions, images, and spatial relations represent more abstract concepts. (This abstractness is one of the reasons that, as noted earlier, it can be hard to tell what exactly someone is talking about based on her gestures alone.)

Third, and finally, gestures can mean because of a convention, much the way that the English word 'chair' means because of a rule that links a particular sound and meaning. This meaning relation is known as *conventionality*. Conventional gestures – like the 'thumbs up,' or the 'shhh gesture' – are called 'emblems.' Like words, they have standards of form.

You cannot change a feature of an emblem's form and expect it to mean the same thing, just as you cannot insert a new syllable at the start of a word and expect people to know what you are talking about. Emblems are what first comes to mind when lay people think of gesture, but they are not necessarily the best studied (but see Ekman and Friesen 1969, Morris et al. 1979). Researchers interested in cognitive aspects of gesture have been drawn to gestures that are idiosyncratic rather than rule-bound in nature, especially iconic gestures and, to a lesser extent, pointing gestures. It is these gestures that are taken to offer a "window onto the mind" (McNeill 1992, Goldin-Meadow 2003).

In some cases, a gesture 'means' in one and only one of the above three ways, which has led the literature to talk about a few main gesture types: 'deictic gestures' (those that involve indicating), 'iconic gestures' (those that look like their referents in some way), 'metaphoric gestures' (those that iconically represent abstract ideas), and 'emblems' (those that are frozen conventions). However, gestures regularly combine indexicality, iconicity, and conventionality to various degrees (McNeill 2005, Enfield 2009). Consider the beckoning gesture you might make when you catch a friend's eye across a room and want her to approach. The gesture is indexical in that it 'points' toward the intended recipient; it is iconic in that it represents the path between the gesturer and the recipient; and, finally, it is conventional to the extent that, in some cultures, people beckon with the palm up (e.g. in the US) and, in other cultures, they beckon with the palm down (e.g. in Mexico). Idiosyncratic gestures lack a conventional component, by definition, but they still often combine iconicity and indexicality (for discussions, see Goodwin 2007, Cooperrider 2014).

If you pay close attention to gestures in everyday conversation, you will notice that a gesture's meaning is sometimes murky – it may not crisply point or depict, and it may not have a clear conventional meaning either. For instance, speakers commonly make movements that are time-locked with speech, but that do not mean in any of the three ways described above. These movements are often called 'beat gestures' because they keep the rhythm of language but do little else (McNeill 1992). The communicative and cognitive functions of beat gestures remain poorly studied and thus somewhat mysterious, but they cannot be swept under the rug. Nor can so-called 'interactive gestures,' another class of gestures that are common but function differently from the gestures just described. Several researchers have proposed a basic split between gestures that operate at the level of what the speaker is talking about – called 'content' gestures – and gestures that operate at the level of the conversational interaction itself – called 'interactive gestures' (Bavelas et al. 1992; see, also, Kendon 2004 for a related class termed 'pragmatic gestures'). Speakers use the latter type of gestures to clarify who has the floor, to engage the audience, to carry out speech acts like requesting, or to

provide commentary on their own utterances. Here we follow the field's focus, which has been on idiosyncratic gestures that relate to the *content* of what the speaker is talking about – the events, characters, relationships, and ideas that form the substance of our stories and explanations. However, we also note that gestures on the fringes of this traditional focus – including emblems, beat gestures, and interactive gestures – are promising ground for future researchers to chart.

8.3   Gesture and Language

One of the foundational observations of modern research on gesture is that it is tightly linked to language – so tightly that it is sometimes argued that gesture should be considered part of language. The tight linkage between gesture and language is evident across different levels of analysis and across timescales. We first review some of the evidence for this integration, with a focus on how speech and gesture go hand in hand in the moment-to-moment flow of discourse (thus leaving aside interesting issues about how gesture and language develop together in children, e.g. Goldin-Meadow 2014, and about how they may have evolved, e.g. McNeill 2012). We then turn to the ways in which gestural meaning is able to pick up where linguistic meaning leaves off.

8.3.1   Gesture and Language, Hand in Hand

A first way that gesture and language go hand in hand is in their timing. Gestures do sometimes stand alone as utterances in their own right – consider the beckoning gesture described above, or a headshake in response to a yes–no question. And gestures also sometimes fill gaps in the speech (for discussion of these cases, which are sometimes called 'component gestures,' see Clark 1996). More commonly, though, gestures and speech overlap and are finely coordinated in how they overlap. Invariably, the stroke of the gesture will line up in time with whatever part of the speech stream it is most closely related to in meaning. This synchrony is referred to as *co-timing*, and it has an interesting signature: Gesture strokes often precede the relevant unit of speech, and sometimes line up with it exactly, but almost never follow it (Schegloff 1984, McNeill 1992, Morrel-Samuels and Krauss 1992).

A second major way that gesture and language go hand in hand is that they co-express meaning. This synchrony is seen, first, in how the spoken and gestural streams together serve to highlight the parts of an utterance that are important. Speakers are selective in choosing what part of the speech stream they stress, and they are also selective in choosing where in an utterance to place a gesture. Interestingly, these choices often coincide: Gesture will most often go along with whatever is stressed in speech

(Kendon 2004). This timing characteristic applies to beat gestures as well as iconic gestures (McClave 1994). In the case of iconics, the part of the speech stream that a gesture relates to in meaning (which is also the part it aligns with in timing) is sometimes called the 'lexical affiliate' (Schegloff 1984, Morrels-Samuels and Krauss 1992). Thus a speaker might say *cut* while depicting the action of using scissors, *circle* while tracing a circular shape with the fingers, or *above* while moving the hand upward. In other cases, the gesture lines up, not with a single word, but with a larger phrase. And, in cases of the 'interactive gestures' mentioned earlier, how speech and gesture align may be governed by other principles (e.g. Kendon 1995). Finally, as we describe in more detail later, there are deviations from this tidy co-expressivity and these deviations turn out to have important cognitive implications.

Relationships between gesture and speech in timing and meaning are strikingly robust. For example, when stutterers start to stutter, they pause in the gesture too, thus preserving the lockstep timing between the two modalities (Mayberry and Jacques 2000). More generally, when speakers experience disfluencies and decide to repair their speech, they tend to reproduce their gesture as well (Seyfeddinipur and Kita 2001). Findings like these lend support to the idea that the co-expressivity of gesture and speech is not an accident – rather, in McNeill's (2005) terms, speech and gesture exhibit an 'unbreakable bond.'

8.3.2 Gestural Meaning Where Linguistic Meaning Leaves Off

Now we come to an important wrinkle. Though gesture and speech may exhibit an unbreakable bond and work together to co-express meaning, they rarely mean in precisely the same way. Then again, how could they? They are two different modes of expression with different properties. Words are fully conventional and are strung together in linear order according to well-defined rules. Gestures are usually created on the spot and communicate meaning in global chunks rather than in an analytic fashion (see McNeill 1992 for discussion). Moreover, words are categorical bits of meaning, whereas gestures are analogue three-dimensional images that unfold in time. Consider an analogy to another case where word and image are called upon to co-express meaning: The titles of works of art (discussed in Enfield 2009). There are cases when a painting has a clear point of focus (e.g. the mountain called Denali) and the title seems to capture it perfectly (e.g. 'A View of Denali'). But even in such seemingly unproblematic cases, the image involves particulars *not* captured by the title: The perspective, the fall of light, the time of year, and so on. Observations like these are, of course, what inspired the saying that a picture is worth 1,000 words. And so it is with gesture. The point is not that gestures are hyper-realistic and language is impoverished; in fact, gestures are often highly schematic. The point is that a number of choices

about perspective, orientation, handshape, and so on, may be evident in gesture even when they are not evident in the accompanying speech.

To make this point concrete, consider the utterance *Then I opened the door*. Co-timed with *opened*, the speaker reaches out a hand, turns an imaginary knob, and pulls the door toward him. No problem – gesture and speech seem to tidily co-express the idea of opening a door. But this tidiness is a bit of an illusion. Think about all of the different ways this gesture could have been produced, each expressing a variant on the idea of opening a door. It could have been a swinging door with no knob, in which case the gesture may have depicted a pushing motion; it could have been a door so heavy that two hands were needed to open it; and so on. This range of possibilities is not just a peculiarity of the word *door* in English – consider aspects of meaning left out in sentences like *The café is next to the dry cleaners* (on which side?), or *It was a chocolate cake* (what size or shape?). Over the years, cognitive linguists have taken much interest in how subtle choices of linguistic form reflect differing *construals* of the same underlying event or situation (e.g. Croft and Cruse 2004, Bergen 2012). This observation can be extended further: Even utterances that are identical in form may involve different underlying construals, and one way we could detect a difference in construal is by attending closely to subtleties of the co-occurring gestures. In the next sections, we examine the insights gesture offers into how people are construing events or ideas by considering several different dimensions of meaning.

Perspective. When you tell a friend that *A rat scurried across the room*, you probably have some specific image in mind. But it is hard to tell just what that image is from your speech alone. For starters, the words alone tell us nothing about the perspective you are taking on the scene. Are you taking the view of the rat, perhaps imagining the quick arm movements that might be involved in scurrying? Or are you taking the vantage point of an observer, imagining the path of the rat's movement across your field of view? These two different possibilities have been observed in gestures and are known as 'character viewpoint' and 'observer viewpoint,' respectively (McNeill 1992, Parrill 2010). Whether the speaker takes a perspective inside or outside of a scene is also evident in the gestures (and, interestingly, the signs) produced when describing spatial configurations. In the case of spatial descriptions, the different viewpoints are called 'route' (inside perspective) and 'survey' (outside perspective) (Emmorey, Tversky, and Taylor 2000). Note that, in both cases, the language used in the broader discourse may tend to resonate with the perspective adopted in gesture. The crucial point is that, for any given utterance, this perspective may be evident only in gesture.

Spatial orientation. Let us consider the scurrying rat in a bit more detail. Another aspect of the scene, one that is easy enough to leave out of speech, is which *direction* the rat was scurrying. The word 'scurry' on its own says nothing about direction – it does not specify whether the rat scurried left

or right, up or down, east or west. But a scurry gesture has to move in one direction or another. Kita and Oyzurek (2003) found that speakers' gestures reliably preserved the direction of a motion event in a recently observed cartoon, even though speakers' words rarely mentioned it (interestingly, in other respects, the same gestures and speech were tightly co-expressive). Similar observations apply to the representation of static scenes, where a direction-free word like 'next' might be co-produced by a direction-specifying gesture (Tutton 2013). In these studies, speakers' gestures reliably capture spatial direction or orientation in terms of the bodily coordinates of the observer; for example, in terms of their left–right or front–back axis. It has been suggested that, in cultures where cardinal direction terms are used pervasively, speakers produce gestures that preserve the original cardinal orientation of previously observed events or configurations in space, rather than the orientation with respect to the viewer's body. Haviland (1993) illustrates examples of this phenomenon in Guugu Yimitthirr, an Aboriginal language of Australia. A storyteller describing how a boat capsized recreated the event – in gesture, but *not* in speech – as happening from east to west. This orientation is clear because the same cardinal orientation was preserved across two different tellings of the story in which the storyteller faced different cardinal directions. Consistent use of cardinal orientation in gesture has also been documented in speakers of Yucatec Mayan when describing the spatial relationship between two static landmarks (Le Guen 2011).

Embodied action. When speakers gesture about actions like opening a door, wrapping a present, or tying a knot, their gestures often incorporate details about *how* the action was performed – for instance, what kinds of grips or movements were involved. Cook and Tanenhaus (2009) had people solve a classic reasoning puzzle, the Tower of Hanoi. Solving the puzzle involves moving a set of differently sized disks from one side of an apparatus to another, while also obeying certain rules. One group of participants solved the puzzle using a physical apparatus with heavy disks and high pegs. The other group solved a digital version of the puzzle by clicking and dragging with a mouse. The question was whether people in these different conditions would gesture differently when later explaining to another participant how to solve the puzzle. Those in the physical apparatus group, compared to those in the digital group, used more grasping handshapes (like the handshapes you would use when actually picking up the disks) and produced gestures with higher trajectories (like the trajectories you would use when having to lift the disks over the high pegs). Importantly, the researchers found no differences in the language used by the two groups.

Problem-solving strategies. From the preceding discussion, it is clear that gesture offers insights – over and above speech – into how people are thinking about events, actions, and spatial relationships in the physical world. But gesture can also offer insights into how people are approaching

much more abstract problems. An example is children's understanding of a tricky concept: conservation of quantity. If you pour all of the water from a tall, skinny glass into a short, wide glass and ask children whether the amount of water has changed, children under the age of seven or eight will often say that it has. These children (called 'non-conservers') are focusing on the fact that the height (or the width) of the water has changed, and not on the crucial fact that these changes are complementary and compensate for one another. But the gestures of these non-conserving children often reveal something more interesting. Some children produce speech that highlights the *height* of the two glasses (e.g. *This one is taller than that one*) while, at the same time, producing gestures that highlight their *width* (e.g. showing a short distance between the thumb and index finger near the tall glass while saying *taller* and a bigger distance near the short glass when saying *shorter*). Children who produce these so-called 'mismatches' between gesture and speech seem to be entertaining two ideas at once. And, interestingly, it is the mismatching children who are particularly likely to learn from subsequent instruction (Church and Goldin-Meadow 1986). The predictive power of children's mismatching gestures has also been found in reasoning about mathematical equivalence (Perry, Church, and Goldin-Meadow 1988) and the balance scale (Pine, Lufkin, and Messer 2004).

In general, gesture speech-mismatches seem to occur at moments of cognitive instability. This phenomenon is seen not only in children, but also in adults learning a task. For example, when explaining how they solved the Tower of Hanoi puzzle – the same disk-and-peg task described earlier – both children and adults sometimes produce gestures that suggest a problem-solving step that is different from the step described in speech. Interestingly, such mismatches are not randomly sprinkled throughout people's explanations, but crop up at precisely those points in the problem where a decision needs to be made between two alternative steps (Garber and Goldin-Meadow 2002). In another study, Alibali et al. (1999) examined adults' gestures as they described algebra problems that they would later have to solve. Some problems involved discrete change. When people's speech and gestures together represented this discreteness, they tended to use a corresponding strategy to solve the problem later. Other problems involved continuous change. Again, when people's speech and gestures both represented this continuousness, they went on to use a corresponding solution strategy. When people only represented the problem as discrete or continuous in speech, with no gestures, this explanation was not as strong a predictor of their eventual solution strategy. And, more interestingly, when they represented different strategies in speech and gesture (continuous in one, discrete in the other), they were just as likely to use the gestured strategy as the spoken strategy. Speech alone is thus not as powerful a predictor of how someone is thinking about a problem as are speech and gesture together.

Metaphor. Reasoning about abstract ideas is not confined to cases where we are presented with difficult problems in the classroom or laboratory. In everyday conversation, people talk about notions like progress, order, and the passage of time; about intricate social scenarios and moral issues; about status and power; and about many other slippery but important abstractions. As cognitive linguists have insightfully shown, discourse surrounding topics of this sort is packed with metaphor (Lakoff and Johnson 1980a, and Sullivan this volume Ch. 24). Less appreciated is the fact that such discourse is also packed with gestures (Cienki and Müller 2008a). As described earlier, some researchers distinguish a separate type of gesture called 'metaphorical gesture' (McNeill 1992), though 'meta-phoricity' may also be thought of as a dimension of gesture (McNeill 2005). Pointing gestures can be metaphorical, and emblematic gestures can have metaphorical aspects. Research on metaphorical gestures to date has yielded at least two important insights about metaphorical thinking. First, just as gestures about concrete topics can provide vivid insights into visuo-spatial and motor aspects of construal that are not part of speech, gestures about *abstract* topics can provide vivid insights into spatial and motor aspects of metaphorical construal that are not part of speech. Second, as observed initially by Cienki (1998), gesture often lets us know that metaphors are afoot when there would be no hint of the metaphor in a verbal transcript.

Insights from gesture have been especially important in deepening our understanding of how people think about time (see Gaby and Sweetser this volume Ch. 39; see also Cooperrider, Núñez, and Sweetser 2014 for a review). Time has long been a favored model system for understanding metaphorical thinking because it is both ineffable and indispensable (Casasanto 2008b). Consider the statement *The meeting took forever*. In an accompanying gesture, the speaker might depict a large swath of time between her hands, as though conceptualizing the duration of the meeting as an elongated object oriented across her body. Such a gesture tells us, first, that the speaker is construing time as linear and oriented left to right. Just as a gesture depicting a scurrying rat must be oriented in space in some way, a gesture depicting a stretch of time must be oriented too. Second, look back at the speech – there is no trace of verbal metaphor. At a very basic level, the gesture thus tells us that the speaker is reasoning metaphorically when we would not know this otherwise.

Speakers do not spatialize time willy-nilly in their gestures. They tend to gesture systematically in ways that fit with a culturally shared model. The most common pattern for gesturing about time in English, for instance, is to show a timeline oriented from left to right. This orientation is surprising given that verbal metaphors in English (*The week ahead* or *Back in the eighties*) exclusively mention the front–back axis (Casasanto and Jasmin 2012). Speakers do gesture about time along the front–back axis, albeit less frequently than along the left–right axis. And, interestingly,

they also sometimes mix together these two metaphors in their gestures, producing, for instance, a gesture with 'tomorrow' that unfolds forward and rightward and is thus simultaneously consistent with both front–back and left–right metaphors (Walker and Cooperrider 2016). Because time gestures are systematic and culturally shared, they can also give us insights into cross-cultural variation in the conceptualization of time, even when clues to this variation are either ambiguous or elusive in the language alone (Núñez and Sweetser 2006b).

Time has been a major focus of research on metaphorical gestures to date, but related insights have also been gleaned from other domains. Enfield (2009) describes a gestural system used by Lao speakers for construing kinship relations spatially. Winter, Perlman, and Matlock (2013) describe some of the metaphorical gestures speakers produce when talking about numbers, including gestures that construe numbers as big, high, or further rightward along an imagined number line. And Marghetis (2015) describes how gestures reflect different ways of metaphorically construing arithmetic, such as collecting together objects or moving along a path.

8.4   Gesture and Communication

In the previous section, we considered how researchers can glean important insights into subtleties of what speakers are thinking by attending to subtleties of how they gesture. But, clearly, speakers are not gesturing as a service to curious researchers. So who is all this gesturing *for* anyway? Is it for the listener's benefit? Is it helpful for the speaker? The 'who is gesture for' question is of long-standing interest, and it remains hotly contested. To preview, there is not one answer to the question: Gesture serves both communicative functions (i.e. for the listener) and cognitive functions (i.e. for the speaker). In this section, we consider some of gesture's communicative functions.

Do listeners glean information from gestures and, if so, what kinds of information? Which aspects of a gesture – such as its form or its relationship to the accompanying language – play into the comprehension process? Despite some early skepticism in the field (e.g. Krauss, Morrel-Samuels, and Colasante 1991), it has now been demonstrated beyond doubt that listeners extract substantive information from gesture (see Kendon 1994 for a review of earlier studies and Hostetter 2011 for a review of recent experimental findings). A compelling kind of evidence for this claim comes from cases where listeners glean information that is altogether absent from the speech. Recall the Tower of Hanoi study by Cook and Tanenhaus (2009) described earlier, in which speakers produced gestures of varying heights to represent moving the disks from one peg to another. Again, this variation in height was not part of what speakers said – it showed up exclusively in their hands. In a later phase of the

same experiment, the researchers looked at what listeners took away from these gestures. They asked the listeners to solve the Tower of Hanoi puzzle themselves afterward on a computer, so that they could precisely track the listener's mouse movements. What they found was that people who saw higher-arcing gestures during the explanations produced higher-arcing mouse movements during their own solutions. Listeners thus extracted information about a gradient – and task-irrelevant – feature of gesture and incorporated it into their subsequent actions.

Another convincing demonstration of this kind of integration is that listeners incorporate the gestures they see into their own subsequent talk and gesture. Cassell, McNeill, and McCullough (1998) had participants view stories told by an actor, who was trained to produce particular gestures–speech combinations. At one point in the story, the actor says *And Granny whacked him one* with an accompanying gesture. There were two versions of this key moment: in one, the accompanying gesture was a punch and, in the other, it was a slap. When retelling the story, a listener who saw the version with the punch gesture reported verbally that Granny *punches him*; a listener who saw the version with the slap used the word *whack* and demonstrated the slap in gesture.

These findings show clearly that listeners can, and do, extract meaning from gesture. Recent studies have gone a bit farther, showing that listeners integrate gesture even when they are instructed to ignore it. Kelly, Ozyürek, and Maris (2010) had people watch 'action primes' – clips of simple actions like chopping vegetables. Immediately after viewing the prime, they were presented with a target video in which a speaker made a gesture and said a word, and the participants had to judge whether the new video related to the earlier one. People were fastest to respond when both speech and gesture related to the earlier video (e.g. the word *chop* with a gesture representing chopping). Interestingly, this effect was found even when participants were told to pay attention *only* to the speech in the target – they could not help but process the gesture as well as the speech. Automatic processing of this kind has also been found for pointing gestures (Langton and Bruce 2000). A new frontier of research is also examining the cognitive and neural processes that underlie gesture–speech integration. For instance, to what extent does understanding gesture engage neural circuits that support spatial reasoning or motor imagery (Ping, Goldin-Meadow, and Beilock 2014, Wu and Coulson 2015)? To what extent does gesture understanding engage classic language circuits, such as Broca's area (Skipper et al. 2009)?

Listeners glean information from gesture, but does their doing so have real-world consequences? Yes, it does. The best evidence comes from studies of gesture's role in the classroom. Consider cases such as those described above where children produce gesture–speech 'mismatches' when struggling to explain a new concept like conservation of quantity or mathematical equivalence. This information is not

just useful for researchers trying to predict who will learn, but is also accessible to teachers and ordinary listeners, and information found uniquely in teachers' gestures is accessible to their pupils (see Goldin-Meadow and Alibali 2013 for a recent review). Gestures that learners see during instruction can thus play a role in helping them form new concepts. Lessons about mathematical equivalence (Church, Ayman-Nolley, and Mahootian 2004), symmetry (Valenzeno, Alibali, and Klatzky 2003), and conservation of quantity (Ping and Goldin-Meadow 2008) have been found to be more effective when they include gesture than when they do not. In fact, lessons may be particularly effective when they include gestures that provide information over and above the information that is in speech. For instance, Singer and Goldin-Meadow (2005) instructed children in how to solve mathematical equivalence problems using explanations in which either (a) the *same* strategy was presented in speech and gesture, or (b) two different strategies were presented – one in speech and one in gesture. Children receiving two different strategies outperformed those receiving one. They even outperformed children receiving the same two strategies but conveyed entirely in speech. This result is striking in that not only were children able to make sense of the mismatch between speech and gesture, but they also benefited from it.

The kinds of evidence described in the preceding paragraphs show clearly that gesture shapes what listeners understand, and that this shaping has consequences later on. An important related question is whether speakers *intend* their gestures to communicate in the first place. After all, gestures could be largely involuntary, perhaps in the same way that facial expressions have been argued to be involuntary (Ekman and Friesen 1969). In some cases, it is clear that speakers are tailoring their gestures to communicate. Pointing gestures co-produced with *this* or *that*, or iconic gestures co-produced with *like this*, are good examples of such tailoring. And, in fact, when speakers cannot see their listeners, they do not call attention to their gestures verbally (Emmorey and Casey 2001, Bangerter 2004). But even gestures that are not explicitly marked in this way may be intended to communicate. Evidence for this possibility comes primarily from studies in which speakers cannot see their listeners – perhaps because they are talking through a wall, on the telephone, or over an intercom. Studies of this sort consistently find that speakers gesture less when their listeners cannot see them (though, interestingly, production of beat gestures is not affected) (Alibali, Heath, and Myers 2001, Bavelas et al. 2008). Notice, however, that these findings have an intriguing flipside – they make it clear that communication cannot be the *whole* story, as speakers continue to gesture at non-negligible rates even when their listeners cannot see them. This is a first hint that speakers may have 'selfish' motives after all.

8.5 Gesture and Cognition

We may not have needed an experiment to convince us that people gesture when there is no listener to see them – this behavior is easy enough to observe in ourselves when we talk on the phone. Other cases of 'gesture for self' may be a bit more surprising. People continue to gesture in completely non-communicative contexts, such as when alone solving problems involving mental rotation (e.g. Chu and Kita 2011) or counting objects (Carlson et al. 2007). Congenitally blind people gesture even though they have never seen a gesture in their lives, and are talking to other blind people who have not, either (Iverson and Goldin-Meadow 1998). Even in cases where certain kinds of gesture are taboo – as is the case of pointing with the left hand in Ghana – people simply cannot help themselves (Kita and Essegeby 2001). In short, we all seem to experience what Kita and Essegeby (2001) describe as a 'cognitive urge' to gesture.

One reason for this urge may be that gesture helps us talk and think. For example, gesturing can help us access the right words (Krauss 1998), particularly in talk about *spatial* topics. Speakers produce gestures at higher rates when talking about space than when talking about non-spatial topics. Further, when people are prevented from gesturing by physically restraining their hands, this manipulation selectively disrupts their ability to talk fluently about space (Rauscher, Krauss, and Chen 1996). Another idea, not incompatible with the possibility that gesturing aids lexical access, is that gestures help speakers organize their thinking in a way that is suitable for verbal expression (Kita 2000).

Gesture may also help thinking by lightening 'cognitive load.' People can only handle so many cognitive tasks at once – handling several at the same time leads to greater cognitive load. Gesturing while speaking could be thought of as a cognitive task that increases cognitive load. If so, gesturing should hinder our ability to handle other tasks performed at the same time. But if gesturing forms a single integrated system with speaking, then gesturing could actually free up cognitive resources that can be used to perform other tasks. To examine these possibilities, Goldin-Meadow et al. (2001) gave people two tasks to perform simultaneously – one was a memory task, the other an explanation task. Would telling people to gesture during the explanation task help or hinder performance on the memory task? As it turned out, people did *better* on the memory task when they gestured than when they did not gesture. In this experiment, participants' gestures naturally related to what they were explaining. It has since been found that 'meaningless movements' – rhythmic movements that do not relate to the content of the speech – do *not* lighten the speakers' cognitive load in the same way (Cook, Yip, and Goldin-Meadow

2012). Movements need to be meaningfully integrated with speech in order for them to lighten cognitive load.

In addition to helping speakers talk and lightening their cognitive load, gesture can also play a role in reasoning. Alibali et al. (2011) showed that people's gestures may lead them down particular reasoning paths when solving gear problems. In this type of problem, people are shown a configuration of multiple interlocking gears, and are asked how moving one gear would affect another gear in the configuration. People typically use one of two strategies: (1) a 'depictive strategy' in which they simulate the physical movement of each gear involved; or (2) an 'analytic strategy' in which they determine whether the number of interlocking gears is odd or even. Encouraging people to gesture in this task nudges them toward using the depictive strategy, presumably because gesture lends itself to visuo-spatial simulation. Relatedly, Beilock and Goldin-Meadow (2010) have found, using the Tower of Hanoi puzzle discussed earlier, that the gestures speakers produce when explaining how they solve the puzzle have repercussions when they try to solve similar versions of the puzzle later on.

Some of the best evidence for gesture's cognitive functions comes from studies of the role of gesture in learning. These studies also make it clear that gesture can have important real-world consequences for the gesturer. As we have already seen, the gestures that learners produce when encountering a tricky new concept can often tell us something about their cognitive state. But, it turns out, those gestures can also be instrumental in changing that state. The mere act of encouraging children to gesture – without telling them how – appears to activate their unspoken ideas, which, in turn, prepares them to benefit from later instruction. This kind of benefit has been found in mathematical equivalence (Broaders et al. 2007), as well as in moral reasoning (Beaudoin-Ryan and Goldin-Meadow 2014). Moreover, teaching children to produce *particular* types of gestures can give them totally new ideas that they are able to transfer to new problems (Goldin-Meadow, Cook, and Mitchell 2009). Even subtle properties of the gestures learners are asked to produce can make a big difference. One recent study compared the learning benefits of two kinds of gestures. One kind represented concrete actions, such as picking up and moving number magnets in order to solve an equivalence problem; the other kind of gesture abstracted away from these concrete actions, instead highlighting an abstract operation – grouping – involved in solving the problem. Children taught to use the abstract gesture developed a more abstract understanding of equivalence, as evidenced by the fact that they were better able to generalize what they learned to new types of equivalence problems than children taught to use the concrete gesture (Novack et al. 2014).

In this and the previous section, we have reviewed evidence that gesture serves both communicative and cognitive functions – that it is

both for the listener and for the speaker. We have also described how, even though it may not be obvious in the moment, the gestures we produce have consequences for reasoning and understanding. Importantly, the fact that gesture can serve all these functions does not mean that it is always serving those functions in each and every instance. A challenge for future research is to work out in better detail *when* gesture serves communication and *when* it serves cognition, as well as when it has important cognitive consequences. Not all gestures are alike, and different gestures may turn out to have different functions and consequences.

8.6 New Directions and Conclusion

Our focus has been on relationships between spoken language and gesture as they are co-produced in the moment-to-moment flow of discourse. A number of additional aspects of the language–gesture relationship would no doubt also be of interest to cognitive linguists. Several of these have already received substantial attention, including how gesture and language relate during language development (see Goldin-Meadow 2014, and Goldin-Meadow and Alibali, 2013 for a review); how gesture relates to language when it is produced in the manual modality (i.e. sign language; see Goldin-Meadow et al. 2012, Goldin-Meadow and Brentari in press); and how gesture varies across cultures, in ways that do and do not parallel variation in language (see Kita 2009). Other topics remain to be explored in fuller detail. For example, does gesture lead to 'framing effects' like those that have been described for spoken language (Thibodeau and Boroditsky 2011)? To what extent are gestural forms conventionalized, and how does such conventionalization occur? How do the communicative and cognitive functions of gesture compare to those of its close cousins – for example, the words of spoken language, diagrams, facial expressions, and the signs of sign language? What is the role of gesture in the creation of new communication systems, and what can this tell us about its possible role in language evolution (Fay et al. 2014)? And many more.

For centuries, language has been a topic of singular interest among researchers working on human communication and cognition. Much of this work has focused on the formal side of language – its properties as a complex system of words and rules. But there is another side of language hiding in plain sight – *gesture*. Gestures show that, when people talk, they are doing more than using the words and rules of a formal system. Gestures and speech are intimately interwoven even as, at the same time, they are distinct – gestures are imagistic and often idiosyncratic, whereas words are categorical and conventional. This synchronization does not mean, however, that gesture and speech are translation

equivalents in different modalities. Gestural meaning often picks up where linguistic meaning leaves off, affording the researcher vivid insights into how speakers are construing what they are talking about. Gesture is much more than just a tool for researchers, however. It also serves the listener – whether learner, teacher, or lay observer – by adding layers of meaning to spoken messages, and it serves the speaker by aiding talking and thinking and by shaping reasoning.