8. THE THOUGHT-LANGUAGE-HAND LINK AND LANGUAGE ORIGINS

INTRODUCTION

This final chapter is framed by the following admittedly broad question: Why is imagery part of language? That it is so has been a leading theme of this book. I shall develop a hypothesis at the end of the chapter that the evolution of language crucially depended at one point on gestures and imagery. Without gestures, according to the hypothesis, some of the brain circuits required for language (Broca’s area, including ‘mirror neurons’, and possibly others) could not have evolved in the way they have, whereby thought and language orchestrates them directly via a thought-language-hand (and speech) link. This hypothesis entails that the integration of gesture with language that we observe today in ourselves, and which in this book we have been at pains to describe, is an essential part of the machinery that was selected in evolution. It further entails that gesture is not a behavioral fossil, not an ‘attachment’ to language or an ‘enhancement’, as the theory that the first form of language was gestural implies, but is an indispensable part of our current day ongoing system of language, and was selected with speech in the evolution of this system.

THE THOUGHT-LANGUAGE-HAND LINK

A claim of the GP hypothesis is that gestures and instrumental actions should be dissociable. Gestures are generated as integral components of the activity of speaking. Actions, such as picking up objects, are not integral components of speaking - they may co-occur with it, be described by it, etc., but are not actually part of it. Against this claim, one could argue that meanings activate actions, and these actions run normally. The result is a gesture, but the brain circuits are those of action. Although action and gesture cannot be dissociated in unimpaired speakers, if the implication is correct, a speaker whose physical condition prevents the normal occurrence of instrumental actions, but who is not paralyzed and whose cognitive and linguistic faculties are normal, should still produce appropriately formed speech-synchronized gestures. This is because gestures arise from the processes of thinking and speaking, and can arise separately (at least in part) from the brain systems controlling instrumental actions.

The man referred to as IW is such a speaker. His case shows a dissociation of gesture from instrumental action, and suggests the existence in the human brain of a dedicated thought-language-hand link that is separate in part from the paths of action control. Such a dedicated link, used only for language, could have arisen as part of the origin of language.

Relationship to the brain model

Broca’s area is the most likely place for a dedicated thought-language-hand link. The brain model described in Chapter 7 focuses brain processes underlying GPs and unpacking on the left anterior speech area (Areas 44 and 45, or Broca’s area). It is here that inputs from the right hemisphere, the frontal areas, and the posterior regions of the left hemisphere converge. It

1 A version of this chapter appeared in CLS 39.
is also the part of the brain that appears to be specialized, as in all primate brains, for the orchestration of movements under some significance, as described in Chapter 7.

IW CASE – NEUROPATHY

The man known as IW suffered a sudden loss, at age 19, of all sense of touch and proprioception (movement and position sense) below the neck, due to an infection of unknown etiology. The immediate cause of damage is thought to have been an auto-destructive immune reaction. The damage was proportional to the extent of mylenation of the sensory fibers, more damage occurred with more mylenation. The slow-conducting, unmylenated sensory nerves concerned with pain, temperature and such ill-defined modalities as muscle fatigue were unaffected. Likewise the motor nerve fibers were left intact. But IW’s fast-conducting, mylenated proprioception and spatial position sense fibers were totally destroyed. Although not paralyzed, IW lost all motor control that depends on these sources of feedback and was unable to walk, write, or feed himself. Speech and all cognitive functions were completely unaffected. The medical providers at the time gave him the grim prognosis that he would be wheelchair-bound for the rest of his life. But after some 30 years of effort IW has taught himself to move again using cognition and vision in place of proprioception and the spatial position sense. He has been so successful that to a causal observer, nothing appears out of the ordinary. He moves his arms and hands accurately and with speed, is able to stand up, walk and sit down with balance and fluidity, and can track his shifting center of gravity as he leans over to pick something up from a low table. His effort and motivation are succinctly captured by the title of Cole’s 1995 book, Pride and a Daily Marathon. Without vision, IW has no idea of where his hands are located in space. To imagine what deafferentation might be like, try this experiment devised by Shawn Gallagher:

- Sit at a table with your hand placed underneath it, out of view.
- Make a fist and then extend your index finger.
- Curl it back into the fist and then extend it again.
- Move one hand below the other.

The mechanisms that allow you to tell when one hand is below the other or your finger is extended or not, or even that you've made a fist, simply do not work for IW.

IW’s self-descriptions of his motion and motion control are lucid and clearly articulated. I take these descriptions as important clues in the discussion that follows. He considers that he has discovered new ways to access, activate, and monitor old motor schemas that he (as anyone) had developed over the first years of life. These schemas had received feedback from the proprioception and spatial position senses and then suddenly had been cut off from them, but the schemas were still embedded in motor memory and over time were accessed in new ways. In his description, he distinguishes two kinds of gestures – in his words, ‘constructed’ gestures versus

---

2 The study of IW is ongoing and is being carried out with Jonathan Cole and Shaun Gallagher. I am drawing on Jonathan Cole for this background sketch of IW’s neuropathy (Cole 1995). I am greatly indebted to IW himself for his willingness to devote time, effort and travel over long distances to the project of studying his gestures and language.
‘throw-away’ gestures. To IW the ‘constructeds’ are the most important: he makes them in order to add naturalness to his speech as part of a deliberate communicative strategy in which information is divided into gestured and spoken streams. For us, the gesture-observers, the ‘throw-ways’ might be the more interesting. They are apparently natural gestures, unintended, unwitting, unmonitored, and occurring spontaneously with speech. They show all the hallmarks of naturally occurring, spontaneous gestures from speakers lacking IW’s injuries – synchronous with co-expressive speech, out of awareness and not monitored visually, with no gaze diversion even when vision is available.

What’s interesting?

We can investigate IW’s movements in terms of two different categories of motion: (a) instrumental actions and (b) speech-accompanying gestures, and this can be done under two conditions – with and without vision. If IW is unable to control instrumental movements when vision is denied, but can continue to perform gestures with full accuracy and synchrony with speech under these same conditions, we will have demonstrated that gesture can be dissociated from action. Such dissociation suggests the existence of a specific thought-language-hand linkage in the brain, distinct at some point from the pathways for controlling actions. This pathway could be non-dependent on proprioception and vision as sources of control, if it is primarily organized by the system of meaning that is still accessible to IW.

THE EXPERIMENTS

The BBC brought IW, Shawn Gallagher and Jonathan Cole to the University of Chicago for filming in July 1997.3 Our aim was to record IW under a variety of conditions, both with and without vision. IW cannot be simply blindfolded, since he would be deprived of the visual input which enables him to orient himself vertically, and would be at serious risk of falling over. We devised a tray-like blind, pictured in Figure 8.1, that could be pulled down in front of him at roughly neck level, blocking his vision of his hands while allowing him room to move and preserving his visual contact with his surroundings.4 IW was videotaped retelling the standard animated cartoon stimulus. He also was recorded under the blind in a casual conversation with Cole.5 In 1997, we did not appreciate the importance of testing IW’s instrumental actions without vision but we had an opportunity to test his performance on this kind of task in April 2002, when IW and Jonathan Cole came back for a second visit to the University of Chicago.6

3 The film was broadcast as part of the BBC Horizon Series, with the title, “The Man Who Lost His Body” (1998).
4 Nobuhiro Furuyama suggested the blind experiment. David Klein built the blind itself.
5 We also had IW wear an eye-tracking device as he described the routes he had followed that morning from his hotel to the psychology building and within the psychology building itself, while looking at maps and floor plans, but I will not describe these results here.
6 The second round of experiments was supported by a grant from the Wellcome Trust to Jonathan Cole.
Example gesture (2002)

The accompanying illustrations (Figure 8.2) show one of IW’s gestures with vision during a conversation, videotaped in 2002. The panels show three successive phases of a single gesture. The co-expressive speech was “gesture for me falls into a number of areas”, and the gesture (about a half second long) displays a metaphorical image of a sorting process. The full gesture included three repetitions of the circling motion. His gaze, as we can observe, was directed downward at his hands. He is explaining that he distinguishes two kinds of gestures - in his words, ‘constructed’ gestures versus ‘throw-away’ gestures.

Fig. 8.2. IW metaphoric gesture of sorting with vision.

Significant variables in assessing IW’s gesture performance

To have a systematic approach to IW’s gestures, we pay specific attention to the following variables (all but the second and third have figured in earlier chapters):

- **Timing**: synchronization with co-expressive speech.
- **Morphokinesis**: this refers to the shape of the gesture in terms of hand forms and use of space.
- **Topokinesis**: this, in contrast, refers to the location of hands relative to each other in space, including but not limited to the approach of one hand by the other.
- **Character viewpoint (CVPT)**: gesture made from the perspective of the character being described. Such a gesture is close to mimicry.
- **Observer viewpoint (OVPT)**: gesture made from the perspective of the narrator or an observer.

With vision, IW’s gestures collectively display the above features. Without vision, they show some but not all of them: gestures without vision have exact timing with speech, morphokinetic accuracy, and OVPT, but not topokinetic accuracy and CVPT becomes rare. The loss or reduction of these two particular features implies that the gestures performed without vision are not replications of actions: at some point IW’s gestures without vision depart the pathway of action control. The preservation of speech-gesture synchrony however implies that the system which remains is integrated with speech. It is this ensemble of preserved and lost features that points to a dedicated thought-language-hand link.
IW’s gestures with and without vision (1997)

IW’s gestures with vision are similar to those produced by normal speakers, but are fewer in number and tend to be isolated. That is, they are performed one by one in keeping with a constructed-gestures procedure. The next illustration (Figure 8.3) shows one of these gestures made with vision. IW was describing Sylvester inside the pipe after he had swallowed the bowling ball. Both morphokinesis and topokinesis are indistinguishable from that of normal speakers. The hand appears to bracket a small figure in the central gesture space and it move downward, wobbling it right to left. The motion is co-expressive with the synchronous speech: //he // wiggles his way down. Note that IW looks at his hand during the stroke. The viewpoint in this case is that of an observer (an observer who is moving an object: not enacting a character, Sylvester), although in his full description of the bowling ball episode the character viewpoint also occurs:

OVPT: “tiny little bird”
CVPT: “bowling ball”
OVPT: “wiggles his way down”
CVPT: “places it”
OVPT: “gets a strike”

The next illustration (Figure 8.4) is a gesture without vision. The gesture was a coordinated two-handed tableau in which Sylvester is represented by the left hand and a trolley pursuing him by the right. IW was saying, “[and the 1 tram 2 caught him up]” (numbers referring to the locus of the first and second panels of the illustration respectively).

His right hand moved to the left (catching up with the left hand) in exact synchrony with the co-expressive “caught him up”. Moreover, a poststroke hold extended the stroke image and maintained full synchrony of the stroke with the co-expressive speech. It is important to recall...
that this synchrony and co-expressivity were achieved without proprioceptive or spatial feedback of any kind. Meaning alone drives the coordination of the hands.

**Topokinetic versus morphokinetic accuracy**

This blind gesture was clearly accurate morphokinetically. It was also noteworthy for IW’s coordinated use of his right and left hands. But - and this is the next important point - it was not accurate topokinetically; the right and left hands did not line up as the right hand approached the left.

The next example (Figure 8.5) illustrates another case of approximation. IW was describing “a square plank of wood” and sketched a square in gesture space. The illustration shows the misalignment of his hands as he completes the top of the square and is about to move both hands downward.

We also asked IW to sketch simple geometric shapes in the air without vision. Morphokinetically a square or a circle was readily created but topokinetically there was always some disparity as the hands sketched the shape (Figures 8.6.1-2, respectively). For comparison, we also asked undergraduate students to sketch geometric figures without vision. In the illustration a student is about to complete sketching a triangle, and is shown at the moment of bringing his right and left hands together to complete the flat bottom. Having topokinetic control, the positioning of the hands is accurate to the millimeter (Figure 8.7).
Similarly, instrumental actions require topokinetic accuracy, and prove to be impossible for IW when vision is denied. The next illustrations (Figure 8.8.1-2) from the 2002 session show two steps in IW’s attempt to remove the cap of a thermos bottle. The first is immediately after Jonathan Cole placed the thermos in IW’s right hand and placed his left hand on the cap; the second is 1 second later when IW has begun to twist the cap off. As can be seen, his left hand has fallen off and is turning in midair. Similar disconnects occurred with other instrumental actions without vision (threading a cloth through a ring, hitting a toy xylophone, etc. – this last being of interest since IW could use acoustic feedback or its absence to know when his hand had drifted off target, but still he could not perform the action).

Significance of the IW results so far

The important implication is that gestures are carried by circuits available to IW without vision, bypassing those required for instrumental actions, namely, those requiring proprioception and spatial position sense of where the hands are placed.

Morphokinetic shaping demands guidance by the meaning of the action but does not demand topokinetic control. The use of space in IW’s gestures under the blind is especially informative. Although he has no exact sense of where in space his hands are, he can align them morphokinetically to create a ‘square’ for example, and this squareness is a direct mapping of the concept of a square onto his motion in space.
The morphokinetic/topokinetic distinction also explains the near disappearance of CVPT without vision. Utterances like “holding it” and “places it” with CVPT gestures resemble Tweety’s instrumental actions of holding the bowling ball and placing it. These CVPT gestures have meaning as simulated actions of the kind for which IW requires visual guidance. It is thus not surprising that CVPT gestures tend to disappear when vision is not available to him.

IW can control speed of speech and gesture output together (1997)

IW can modulate the speed at which, without vision, a meaning is presented jointly in speech and gesture. During a conversation with Jonathan Cole while still under the blind IW reduced his speech rate by about half, yet speech and gesture remained in synchrony (all gestures were metaphoric with metapragmatic content and appeared to be unplanned ‘throw-aways’):

Normal: “and [I’m start]jing to use…”
Slow: “[I’m start]jing to get into… trying to explain things”

Notice in the illustrations (Figure 8.9) how each phase of motion links to speech in the same way at the two rates (IW rotates his hands out of phase at the slow speed and in phase at high speed, but his inward and outward motions occurred at the same points of speech at both speeds).
The rate-change example seems a snapshot of how IW controls gesture movements in the absence of all sensory information. He is able to coordinate his hands and vocal output and change the tempo of them in tandem, because his hand movements and speech are being orchestrated online by a single set of meanings. In terms of the dialectic model, he senses that this GP (with the meaning roughly of starting to do something) is active for a certain duration.
Since the GP is a single process of speech and gesture, any change of its duration automatically affects the two components jointly.

**Summary of IW's gestures without vision**

The following points summarize what we have seen of IW's gestures in the absence of visual, proprioceptive or spatial position sense feedback:

- Gestures preserve synchrony with speech.
- Gestures rarely have CVPT.
- Most appear to be of the constructed type.
- Gestures preserve morphokinetic accuracy, lose topokinetic accuracy.

Hence, gestures are co-expressive and synchronized with speech.

**Phantom limb gestures**


Dr: “How do you know that you have phantom limbs?”  M: “Well, because as I’m talking to you, they are gesticulating. They point to objects when I point to things.”

“...When I walk, doctor, my phantom arms don’t swing like normal arms, like your arms. They stay frozen on the side like this” (her stumps hanging straight down). “But when I talk, my phantoms gesticulate. In fact, they’re moving now as I speak.”

Mirabelle’s case points to a conclusion similar to IW’s – a dissociation of gesture from practical actions, and gestures that occur where instrumental actions are impossible, though obviously for different reasons. In Mirabelle’s case, moreover, GPs create the sensation of gestures when no motion is possible. Presumably, again, the same brain regions (Broca’s area) are the locus of this sensation.

**Overall significance of the IW case**

The IW case suggests that control of the hands and of the relevant motorneurons is possible directly from the thought-linguistic system. Without vision, IW’s dissociation of gesture, which remains intact, and instrumental action, which is impaired, implies (as Shawn Gallagher has said) that the “know-how” of gesture is not the same as the “know-how” of instrumental movement. In terms of brain function, it implies that producing a gesture cannot be accounted for entirely with the circuits that perform instrumental actions; at some point, the gesture enters a circuit of its own. The brain model locates this dedicated thought-language-hand link in areas 44 and 45 (and these processes equally create in Mirabelle the sensation of gesticulation with phantom limbs). These conclusions have implications for the evolution of language, and this topic will be taken up in the concluding section.
EVOLUTIONARY PRECURSORS

I conclude with an argument that will provide an ‘ultimate answer’ to the question with which we began: why does language have the double character that Wundt and Saussure initially saw, whereby imagery and language are joined in a single process? But first I should consider the skepticism with which this question was once regarded.

The 19th C. ban on the evolution topic

Part of the folklore surrounding the language origins topic is that the Linguistic Society of Paris in 1865 banned all mention of it. This in fact occurred; the specific exclusion was as follows:

Article 11: The Society will accept no communication dealing with either the origin of language or the creation of a universal language.

About the same time, the Philological Society of London in 1873 similarly took a dim view of discussion of the topic:

We shall do more by tracing the historical growth of one single work-a-day tongue, than by filling wastepaper baskets with reams of paper covered with speculations of the origin of all tongues

Both of these historic prohibitions are quoted by Adam Kendon in his own contribution to the origins topic (Kendon 1991). That such rules were thought necessary in the 19th C. is evidence of the abiding interest in the origins question. As Kendon points out, justifying his own effort, we can go much farther now than was possible a century and a half ago. We have a more detailed understanding of language, human evolution, biology, the brain, linguistic performance and – crucially – gesture performance and how it relates to speech than existed in the mid 19th Century. Of course, we can never observe the origin of language, but we are able to constrain its possible scenarios far more tightly than was possible in the 19th Century. I submit the following limited evolution model, which takes into account the brain model described in Chapter 7, the IW case just outlined, and the discussion presented in this book as a whole of the dynamic dimension of language and an imagery-language dialectic, to explain how speech and gesture join together to embody meanings.

The proposed evolution model

The model is meant to explain how the brain became able to combine hand movements and vocal action sequences under some significance. Two things had to occur:

Area 44 (one part of ‘Broca’s area’) was, pre-adaptively, the area of the brain where action sequences of all kinds are organized. For language to evolve, it had to be reconfigured to accept language-thought inputs.

Area 45 (‘mirror neurons’, the other part of Broca’s area) also became self-responding to actions imbued with meaning. This was a new step and implies that this region had been co-opted by the language-thought route during the evolution of language capacity.7

---

7 See Chapter 7 for references documenting the orchestration functions of Broca’s area. In addition, Hamzei et al (2003), using fMRI, find that the peaks of action recognition and production of verbs naming displayed actions
Both steps were critical for bringing gesture and vocalization together, and would have been steps not shared by other creatures closely related to humans otherwise, including other primates with their own mirror neuron systems. In other primates this system is limited to responding to actions (both by the self and others) but does not respond to meanings other than the meanings of actions themselves. The reconfiguring of mirror neurons to respond to significances other than actions was a step with wide implications.

Area 44 is not only part of the so-called speech area but also controls other kinds of orchestrated actions, manual as well as vocal (see Chapter 7). These actions seem to be organized under some significance. Accordingly, we can speculate that all actions that involve sequences or stages of different movements are orchestrated in 44 when they have some significance outside of themselves. Area 44’s domain would include speech movements but also manual movements with the requisite complexity, such as gestures. Some evidence of the importance of 44 for orchestration of both speech and gesture is that Broca’s aphasia, which is due to Broca’s area injury, eliminates both speech and gesture sequences (observation due to Susan Duncan; data from Laura Pedelty thesis).

For example, this patient from Pedelty’s data files (Figure 8.10, this patient was discussed in Chapter 7 as well) displays classic agrammaticism but is able to differentiate newsworthy content in both speech and gesture. She is saying “[down d-] [d-down]” and as she does so, her right hand plunges downward (the figure shows the second gesture). Such synchrony of co-expressive speech and gesture suggests a GP at this point. Thus, differentiation is possible but the patient has great difficulty in orchestrating action sequences of either a spoken or manually gestured kind.

The other component of this evolutionary model, the mirror neuron circuit, has become the subject of much interest and speculation since the discovery of these neurons, much of it focusing on their potential for undergirding a ‘theory of (another’s) mind’. This connection is worth emphasizing, because it links the emergent sense of self, at age 3, which I described in Chapter 6, with the brain processes underlying a dialectic of imagery and language that is the

“share a common functional architecture”. Ferrari et al (2003) discovered, in the monkey, populations of neurons in F5 – considered the homolog of Broca’s area – that fire when significant mouth gestures are displayed, some of which also fire when the monkey makes communicative gestures. Specifically addressing mirror neuron circuits, Wohlschläger et al. (2003) found that the speed with which human subjects detect action onsets in others (presumably engaging mirror neuron circuits) is not different from the speed of detecting the onset of their own similar actions (engaging a lever), and both differ from the speed of detecting the equivalent action onset by a mechanism, implying linked brain mechanisms for self and other’s actions. And Grèzes et al (2003) observed in humans activation of area 44 when subjects imitated gestures and also executed movements in response to objects. All of these recent studies support the claim that Broca’s area has the potential to orchestrate actions and, in humans, has also been configured to accept inputs that have a significance outside the action itself. I am grateful to Ana Solodkin of the Department of Neurology, University of Chicago, for drawing my attention to these valuable references.
focus of the current evolution model. Mirror neurons provide a basis for recognizing the actions of others. They are a special class of neurons—identified in monkeys (Rizzolatti & Arbib 1998) and assumed to exist in humans on the basis of a goodly amount of evidence (e.g., Decety et al. 1997, Grafton 1996, Rizzolatti et al. 1996) — that fire when specific actions are observed in others. The same neurons fire when a monkey performs the action itself and also when a monkey sees these actions performed. In humans, BA 45 is active when a subject observes an action (Decety et al. 1997, Grafton 1996, Rizzolatti et al. 1996), and BA 44 and 45 are active when he or she performs this action (Goldenberg 1999, Rizzolatti et al. 1999). On the other hand, this system is not engaged when the subject just observes the objects used in the actions, or passive hands (Nishitani & Hari 2000). In other words, this system seems to be a circuit for recognizing intentional goal-directed actions of one’s own or of others.

My argument in applying the mirror neuron discovery draws on Rizzolatti & Arbib (1998), where they state that the development of the human speech circuit (Broca’s-Wernicke’s areas) was a consequence of the fact that the precursor of Broca’s area was endowed, before speech appearance, with a mechanism for recognizing the actions made by others; these are the mirror neurons. This precursor relationship explains why language and gesture are centered in the same brain areas and why this is the anterior left hemisphere. In the monkey, mirror neurons are found in the prefrontal region known as area F5. F5 is generally agreed to be the homolog of Broca’s Area, and specifically of BA 44 in the human brain. Rizzolatti & Arbib cite PET studies which show that subjects had activation in 45 while observing grasping. So the evolution of speech and manual action sensitivity could be linked in Rizzolatti & Arbib’s view.

However, I do not follow them in their attempt to link these precursors directly to high-level grammatico-semantic categories (Rizzolatti & Arbib attempt to deduce case grammar). It appears to me that this bypasses more immediate steps that would have taken place in the origin of language. What would a more immediate effect have been? My proposal is that a new way to organize sequences of movements in Areas 45 and 44 evolved out of the mirror neuron circuit. The crucial new step was the co-opting of these areas by language and meaning. The actions that mirror neurons must react to in this co-opting by language are movements with significances other than as actions, namely, gestures.

MEAD’S LOOP - GESTURE AND THE SOCIAL DIMENSION

In Mind, Self and Society G.H. Mead (1974) articulated the logic of this evolutionary step in a way that brings in the social basis of human language and ties it to gesture. He wrote in what he termed ‘the philosophy of the gesture’:

“‘Gestures become significant symbols when they implicitly arouse in an individual making them the same response which they explicitly arouse in other individuals.’” (p. 47).

In other words, meaningfulness depends on simulating a social response - the social response of another in yourself, a kind of auto-socialization. It is a mechanism seen in its purest form in gesture. This is because linguistic symbols have their own independent sources of meaning that do not depend on this loop.
**Mirror Neurons Complete Mead’s Loop**

Mirror neurons could be the mechanism of a gestural self-response. We can posit a self-response via one’s own mirror neurons and hypothesize that part of human evolution was to have mirror neurons participate in one’s own gesture imagery. It was a capacity that was selected in the course of human evolution. It hooks evolution into Mead’s loop – one’s own gestures can activate the part of the brain that responds to intentional actions, including gestures, by someone else, and thus treats one’s own gesture as a social stimulus.

The critical evolutionary selection then – unique to H. sapiens – was this new self-response by mirror neurons. Mirror neurons complete Mead’s loop in the part of the brain where action sequences are organized – two kinds of sequential actions, speech and gesture, converging, with significance other than motor action itself as the integrating component. Such an evolution brought meanings into the 44/45 areas, where meanings other than meanings of instrumental actions themselves could orchestrate complex, sequential movements. These movements were both manual and oral/pharyngeal/respiratory, orchestrated in this too-narrowly labeled ‘language area’ and sent to the nearby motor areas.

This hypothesis is meant to explain:

The synchronization of gesture with vocalization on the basis of shared meaning other than actions themselves (and this is the way they synchronize).

The co-opting of the brain circuits that orchestrate sequential actions by meanings. These are the circuits exposed in the I/W case.

A further implication is that Mead’s loop treats gestural imagery as a social stimulus. It thus explains why gestures occur preferentially in social contexts of some kind (face-to-face, on the phone, but not when talking to a tape recorder; cf. Cohen (1977).

Co-opting sequential actions by a socially referenced stimulus (imagery) makes a new kind of action possible, action that we call speech and gesture, was an idea first enunciated by Condillac in the 18th Century and developed imaginatively by Vygotsky in the 1930s (Vygotsky 1962). (Another of Condillac’s ideas – that the first form of language was gesture – is discussed later.)

**Homesign revisited**

The homesign phenomenon (Chapter 4.3) can be interpreted in this light as well. The deaf child’s invention of stable recurring gesture-signs; the preference for a certain sequence of signs; the composition of signs out of stable, recurring sign components: all suggest an ability, available to the child merely as a consequence of development, to orchestrate manual actions based on intentions to communicate meanings – movements controlled not as movements, but as signs. Viewing homesigns not as a system of abstract symbols but as patterns of actions is to see the evolution of language as primarily a new method of movement control; this has been our approach throughout. The ‘resilient’ homesign properties, which I listed from Goldin-Meadow (2003b) in Chapter 4.3, can be explained as the co-opting of the motor parts of the brain by language and meaning, a process as available to deaf children as to every member of the species. In the current explanation, homesign properties are guidelines of action orchestration, which is the reason why they are resilient. For example, The locus-actor-action schema that appears in
homesign utterances is an orchestration in which the hand begins at a static spot, moves to the current locus of an entity that itself moves (the actor), and then performs a simulated action (the movement itself). This is orchestration using crescendo, utilizing meaningful elements at each step, and may be a prototypical orchestration schema. It is not an action in itself; it is a thought (‘transformation of an entity’) seizing control of action, orchestrating sequences with a transitive meaning by adding structures in an increasingly motoric direction. The fact that adults, when speech is disallowed, spontaneously adopt the same schema confirms its hyper-availability (Goldin-Meadow et al 1996). Although orchestration may depend on ‘shareability’ (Freyd 1983), these orchestrated movements occur in the absence of conventions, since adults do not adopt them, and this means they are free to expose the predispositions of Mead’s Loop. A conventional language such as English, where the actor appears before the action, requires the addition of convention to maintain a contrary orchestration schema.

Mead’s Loop, via mirror neurons, connects the intrapsychic to the interpsychic

Vygotsky, of course, also provided the insight that the source of internal mental life is the social fabric of the individual’s ongoing participation in the world. Everything, he famously said in his unanswered debate with Piaget, appears in development twice, first on the social plane, later on the mental. The effect of Mead’s Loop is to treat gesture as a social stimulus. Thus it is a bridge for this two-step development. The social stimulus, per hypothesis, is the road over which thought and language co-opt the motor orchestration areas of the brain, making speech and the course of thinking tied to speech itself possible. Thus the Vygotsky insight: with every linguistic act there is the social simulation of Mead’s Loop, whereby the speaker’s own response is like the response of another person.

THE SCENARIO

To complete this model, I propose a sketch of what conditions of life and biological evolution could have selected the role played by Mead’s loop described above – mediating and co-opting areas 44/45 to create the orchestration of motion by language, hence creating the thought-language-hand link. Such a scenario would have been a rich social context of speaking in order for Mead’s loop to have been of adaptive value, and I suggest that family life was this setting par excellence. Before presenting the scenario, however, I will consider a preliminary step, which is the possibility that a neurogestural joining of manual and vocal movements itself was not part of this evolution, that such a linkage was a preadaptation inherited as part of the general primate line of descent.8

Vocalizations and hand movements preadapted?

Research by primatologist William Hopkins (presented at the Evolution, Cognition and Development conference, 2002) reveals that chimpanzees co-produce vocalizations and gestures. The evidence for this is that hand preference (the right hand) is markedly stronger when gestures co-occur with vocalizations than not. The gestures in question are impromptu creations of the moment, with seemingly an indexical function. Stereotypic begging, in contrast, is performed silently with no particular hand preference. Thus, linkage of spontaneous gesture movements

8 Such a preadaptation, tying gesture to speech before the appearance of language, would be a (final?) nail in the coffin of the gesture-first theory.
and vocal tract activity (including breathing\(^9\)) may have been a preadaptation available to the early hominid creatures in which the selection of Mead’s loop occurred.

Hopkins’ chimp, Beleka, hoping to induce one of the researchers to hand her a banana, frustratingly out of reach in front of her cage, extends her right hand on its side with a hoarse panting sound (Figure 8.11). She performs this combination of gesture-plus-vocalization several times. It seems pragmatically geared to get the researcher to retrieve the banana. It is not the classic palm up begging gesture, a stereotypic signal that appears in much the same form across several primate species (Figure 8.12). Although the begging gesture was also made with her right hand, it took place without vocalization.

If a preadaptation for language was gesture + vocalization, it would appear that language could have been, from the start, on a different track from gestures like begging. Gesture + vocalization would have been available for selection if, as argued earlier, the combination came to have adaptive value. Ultimately, it reconfigured Broca’s area. Begging is different from gesture + vocalization in not being tied to vocalization and in apparent function (the impromptu gesture + vocalization seems to have had a general social/defictic function of ‘hey look!’ in contrast to the specific ‘provision me!’ function of the stereotypic gesture: deixis is more fundamental a linguistic function than provisioning).

Family life

Based on standard timelines for the emergence of H. sapiens, we can narrow the dates for the Mead loop and its selection to the interval from 2 MYA to approximately 100-200 KYA.

Gesture presupposes habitual bipedalism – this commenced about 5 MYA (post the separation of the hominid line from the great apez line). The process leading eventually to the co-opting of action orchestration by language in areas 44/45 could not have started until this development, but the step itself did not occur until much later.

The origin of symbolism is linked by Wrangham (2001) to the invention of cooking and by Deacon (1997) to the formation of a gender-based division of labor and ‘marriage

\(^9\) Emphasized by Nobuhiro Furuyama.
contracts’, both dated to about 2 MYA. Jolly (1999) objects to this description as overly patriarchal, but, even conceding her criticism, some major changes occurred in family life for the creatures living at this time.

This form of living was itself the product of many other changes in reproduction patterns, female fertility cycles, child rearing, neoteny and other factors – all of which might have been emerging long before the changes posited in areas 44/45.

Such a social-cultural development was accompanied by expansion of the prefrontal cortex (Deacon) as well as (per current proposal) the reconfiguring of areas 44 and 45. To judge from the time course of the prefrontal expansion, the process was completed only 100,000-200,000 years ago.

All of this suggests a scenario of organized family life by bipedal creatures inducing changes in brain configuration and function. The selection pressure for Mead’s loop would be the family circle (not the romanticized ‘man the hunter’ or ‘man the tool maker’) – this setting offering a role for Mead’s loop where one’s own meaningful movements are treated as a social stimulus, especially as an engine of childrearing and infant care (responding one’s own communication from the viewpoint of the one cared for would be a powerful selection pressure).

Recent opinions of the differences between chimpanzee and human evolution focus on the unique-t- humans mode of reproduction, including, especially, the family. The selection of growth points co-opts vocalization-gesture preadaptations of the kind shown by chimpanzees, and pushes vocalization and gesture into area 44 where they fuse into meaning-orchestrated movements. This evolution cannot be tied to a single step (‘mutation’). It took place, in this theory, through many converging steps, over a long period, resulting in the creation of family life. A proposed time line is as follows.

Timeline

The evolution of a thought-language-hand link could not have started before 5 MYA and the emergence of habitual bipedalism in Australopithicus, but this would only have been the beginning. And even earlier there were presumably such preadaptations as an ability to combine vocal and manual gestures, but not yet an ability to orchestrate these movements by meanings other than the meanings of actions themselves.

5 to 2 MYA – Lucy et al and the long reign of Australopithicus – would have seen the emergence of precursors of language, something a still apelike brain would be capable of, such as Bickerton’s protolanguage (Bickerton 1990); for example, ritualized incipient actions that become signs, as described by Adam Kendon in his paper on language origins.

Starting about 2 MYA with the advent of H. habilis and later H. erectus, there commenced the crucial selection of self-responsive mirror neurons and the reconfiguring of areas 44/45, with a growing co-opting of action by language, this emergence being tied to the appearance of a humanlike family life with the host of other factors shaping this major change (including cultural innovations like the domestication of fire and cooking). Recent evidence

---

10 However, again, recent scanning data suggests that the human frontal cortex is no larger than would be expected of an ape brain with its overall size (Semendeferi 2001; cf. Chapter 7, note 10).

Along with this socio-cultural revolution was the expansion (if it occurred) of the forebrain from 2MYA, described by Deacon, and a complete reconfiguring of areas 44/45, including Mead’s loop, into what we now call Broca’s area. This development was an exclusively hominid phenomenon and was completed with H. sapiens about 200-100 KYA (if it is not continuing; cf. Donald 1991).

Thus, proto-language and then language itself emerged over 5 million years (far from a big bang mutation). Meaning-controlled manual and vocal gestures, synthesized, as we currently know them, under meaning as GP dialectics emerged over the last 2 million years. The entire process may have been completed not more than 100 KYA. To put it otherwise, there have been perhaps only 5000 generations since the human brain fully achieved a dedicated thought-language-hand link.11

CONCLUSIONS – EVOLUTION

General

It is not an accident that the gesture and language centers of the brain are side by side. Mirror neurons in area 45 mimic the categorized imagery of the GP, which some evidence suggests is relayed from the right hemisphere (McNeill & Pedelty 1995), and link it to area 44 for precise coordination in a movement plan.

The result is control of action sequences in Broca’s area by meanings other than the meaning of the action itself - the result of Mead’s loop, which is possibly a unique circuit in the humans brain, and linked in its function and development to the appearance of the self as an independent agent.

The evolution of language could accordingly have included a step whereby the action orchestration in area 44 was co-opted by the systems in charge of meaning formulation. However, it is wrong to identify the ‘evolution of language’ with this event or any other single ‘big bang’ mutation. A panoply of events seems to have been required. Without any one of them it is doubtful that language could have appeared - bipedalism, vocal-manual linkage in the brain, gesture, expansion of the frontal cortex, neotony, family life, and, the favorite of this discussion, mirror neurons capable of responding to gestures, as a component of movement orchestration in Broca’s area.12

Mead had a further important concept of the self in two alternating forms, which he termed the ‘I’ and the ‘Me’. This distinction can be related to Mead’s loop and the steps taken by children toward symbol formation outlined in Chapter 6. The loop creates the first kind of self, the active I. Mirror neurons, acting in the standard (non-loop) way, may create the other, the reacting M e (of course, Mead couldn’t have made this connection himself). A self-aware agent (Hurley 1998) appears to combine these selves. The I and M e can also be related to the proposed co-opting of Areas 44 and 45 by meaning and language via Mead’s loop. The co-opting was done by the active self, ‘I’, which was the self leading to language. The two selves can be seen in this way as products of the language capacity in humans.

Of course, this list is not exhaustive either.
The growth point would have been the product of this evolution by making images into the organizational unit of speech; that is, gestures fuel speech, in this reconstruction, and the growth point is the mechanism for doing so.

THE GESTURE-FIRST THEORY

Several authors have recently revived a hypothesis that goes back to Condillac in the 18th Century – that when language first appeared it was gesture (e.g., Donald 1991, Armstrong et al.1995, Corballis 2002). In the Condillacian view, language was first a kind of sign language; the transition to speech took place separately and later. The gesture-first theory has been re-energized in these proposals, and I will discuss it in some detail, concentrating on Corballis' version. His is the most recent and, to many, the most compelling modern presentation of the theory. However, I will argue that the gesture-first theory is not plausible. Seeming support for it is either ambiguous or misleading, and the theory cannot explain a crucial feature of what actually evolved – the dual reality in which language is both imagery and a socially shareable code.

Gesture-first arguments, with comments

Laterlization

A general argument is that speech and gesture are lateralized in the same regions of the brain (typically, the left). Shared brain regions fit the gesture first, speech-evolving-from-it scenario, since speech actions, it claims, derive from manual actions, and so naturally occupy the same brain regions. However, lateralization is ambiguous as evidence. The facts equally fit the hypothesis that what evolved is the ability to combine speech and gesture: not gesture first, but speech-gesture, as a unit. As we have seen, this hypothesis also predicts (if we can ‘predict’ something that took place millions of years ago) that speech and gesture have adjacent or shared regions in the brain. So we need to consult other evidence.

Speech-gesture competition is unequal, with gesture less affected

Corballis cites an experiment by Pierre Feyereisen (1997), which in turn was inspired by an earlier study by Levelt, et al (1985). I will consider the Feyereisen study first, and begin with a quote from Corballis:

“When people are required to gesture and vocalize at the same time – for example, by simultaneously naming a symbol that appears on a screen and giving a learned handshape for that symbol – gesture and vocalization compete with each other. But it appears to be an unequal competition. Speaking is slowed down slightly by the requirement to gesture, but the gestures are not slowed down by the requirement to speak.” (pp. 100-101).

The gestures referred to are pointing movements, directed at events on a computer screen. The speech consisted of two arbitrary syllables (“ti” for the left side, “ta” for the right side). The finding was that saying a syllable retards the onset of pointing by some 40 milliseconds, compared to initiation without speech, and pointing retards the onset of speech by some 90 milliseconds, compared to speech alone (Feyereisen 1997). The greater effect on speech suggests that gesture is relatively invulnerable to interference. However, there are no growth points in this experiment. Neither “ta” nor “ti” categorizes motion; pointing to the left was not a
“ti” (it would be “left” or some other natural language category). In other words, speaking the syllable and pointing at the screen were independent actions. Independence explains the interference between them. With natural speech-gesture combinations, in contrast, interference is not expected; on the contrary: distilled into GPs, speech and gesture combine. Moreover, of the two response modes, speech was arbitrary and novel, and hence, not surprisingly, was the more vulnerable and the more affected by a concurrent second response.

The inspiration of the Feyereisen experiment was Levelt et al. (1985), in which Dutch speaking subjects pointed at a lamp and said either “dit lampje” (‘this lamp’) or “dat lampje” (‘that lamp), depending on the position of the lamp in a 4-lamp array. A crucial difference between the experiments is that, in Levelt, et al., the words “dit” and “dat” are real words that could categorize the pointing movements – meaning, roughly, a lamp close to the origin (“dit”) or one far from the origin (“dat”). These categorial differences, combined with pointing, might have formed GPs. The experiment did not contain a no-speech or no-gesture condition, but instead compared conditions in which there were different numbers of speech and gesture options, and found, equivalent to the Feyereisen result, that more options in one modality slowed down the initiation of a response in the other modality. But is response initiation the right comparison? For GP formation, it is not. More appropriate would be the moments at which the speaker materializes meaning in speech and in gesture; we then ask: do the materializations occur at the same time? We want something like a ‘G-center’ and an ‘S-center’, comparable to a P-center, but for production rather than perception, and determine if these are synchronized (see Morton, et al. 1976 for the concept of a P-center). Levelt, et al. show that movement apex times are close to, but slightly after, voice onset times. Movement apex, in the case of a deictic gesture, is probably a good guess of the ‘G-center’. However, voice onset is unlikely to be the ‘S-center’. More likely is the prosodic peak of “dat” or even of the entire phrase, “dat lampje” (so the slight anticipation of the movement apex by the voice onset could arise because S-centers are underestimated). The motion apex and voice onset times vary depending on experimental conditions but they vary together, showing that speech and motion remain integrated. Considering these possible times, it seems plausible that, in the Levelt, et al. experiment, where GPs might have taken form, GPs held together, there was no speech-gesture interference within the GP, and certainly no hint that the gesture modality was less vulnerable (since the movement apex followed speech onset in all conditions but one).

I conclude that these two experiments – Feyereisen’s and, before it, Levelt, et al.’s – do not suggest that gesture is a ‘fossil’ supplanted by speech during evolution.

The relationship of gesture to speech is enhancement

This final argument is the key to gesture-first. Two more quotes from Corballis will start it off:

None of this is certain, of course. The absence of a meaningful discourse context for pointing makes a GP interpretation marginal, and the motions themselves may have been less deictic gestures (= indicatory movements creating or locating entities in a deictic field) than instrumental actions (reaching in the direction of the lamp). In this case, the motions might have engaged brain circuits for action but not the thought-language-hand link. Details of the method, especially an instruction to make “expansive gestures” and the many repetitions of similar movements in the course of the experiment, could have prompted an action mode of responding. There is no current information clarifying this ambiguity.
“…[U]nless you are fluent in signed language, you will know that it is very difficult to get your message across using gestures alone. …[P]eople do gesture as they speak, and the gestures can certainly help one make a point, as it were, but they are not sufficient by themselves” (pp. 127-128).

“The adaptations necessary for articulate vocalizations may have been selected, not as a replacement for manual gestures, but rather to augment them. … Even today, of course, language is scarcely every purely vocal. Watch any two people conversing, and you will see that their speech is accompanied by manual gestures and facial expressions that enhance meaning” (p. 217).

I take this argument to be the following: Speech and gesture traded positions over time; originally, speech enhanced gesture but, as speech evolved, the balance shifted and gesture became the enhancer of speech. As we have seen, however, the relationship of speech and gesture is not merely enhancement. The key to the problem is the speech-gesture relationship as we observe it today. To say ‘enhancement’ is just the way the gesture-first theory has to put it, but this misstates the actual relationship. To call gesture the ‘enhancement’ and speech the ‘enhanced’ is to miss the merging of speech and gesture, and to impose a distinction that doesn’t belong. The “up through” example has been the stock illustration of gesture merging with speech, and neither speech nor gesture ‘enhances’ the other in this case. The gesture (rising hollowness) embodied the same idea of Sylvester’s ascent on the inside of the pipe as the words (“up through”) but in a different mode, and the synchrony of these unlike modes fuels the imagery-language dialectic. That there is such a double system might seem odd, but if speech-gesture units were selected in evolution, such is what we would expect in current day speakers, the products of this evolution.

Other weaknesses of gesture-first

Gesture-first and growth points

Finally, and most crucially, the gesture-first theory does not explain the creation of growth points. It assumes that speech (partly) replaced gesture. According to the theory, gesture is a fossil or semi-fossil, whose role has been taken over by speech. The gesture-first theory cannot explain the integration of speech and gesture on any level, let alone over the multiple levels that we have seen, or the GP’s incorporation of context. Gesture, it implies, is an accompaniment to speech, but is not part of it, and the phenomenon of an imagery-language dialectic is an unexplained mystery.

Overlooked data

I consider this critique of the gesture-first theory to be logical and general, not due to overlooked specific results. Nonetheless, there are specific results that seem inexplicable if gestures are just ‘enhancing’ speech; these are the demonstrations of unbreakable gesture-speech binding cited earlier in the book; so to repeat from Chapter 2:

1. DAF. Although speech is disrupted with delayed auditory feedback, gestures are still synchronous with speech. If the relationship is enhancement gesture might replace speech in such an experiment, but it is unclear how it remains bound to speech.

2. Clinical stuttering. Mayberry & Jaques (2000) found that the onset of a gesture completely blocks stuttering and, conversely, the onset of stuttering immediately suspends an ongoing gesture. In other words, there is incompatibility of stuttering and gesture. Again, gesture should not be integrally part of speaking if the process is enhancement.

4. Dissociated from action. Finally IW, described earlier, performs gestures with speech under conditions where instrumental actions are not possible, the gestures being orchestrated by speech alone. IW himself describes gestures that seem to fit an enhancement model: these are his deliberate gestures, but they are the ones not synchronizing with the co-expressive speech. His ‘throw-aways’, in contrast, are synchronized with speech, and fit the GP model.

Reasons for speech

Corballis points out advantages to having speech be the system of communication over gestures alone. There is the ability to communicate while manipulating objects (so man the toolmaker can make tools and talk about them at the same time) and to communicate in the dark. Less obviously, speech reduces demands on attention, since interlocutors do not have to look at one another (p. 191). There are also positive reasons for gestures not being structured on the static dimension. Susan Goldin-Meadow, Jenny Singleton and I once proposed that gesture is non-static because it is better than speech for mimicry: gesture has multiple dimensions on which to vary, while speech has only the dimension of time. Thus speech, by default, becomes the repository of the static structure (Goldin-Meadow, et al. 1996), leaving gesture free for the dynamic dimension. All of these arguments seem valid, but do not bear on the difference between the gesture-first and the gesture-and-speech-together theories. Even allowing that speech has such advantages, and gesture has its own role in mimicry, natural selection could still select gesture-speech combinations, via the Mead’s Loop/mirror neurons circuit, as described earlier.

CONCLUSIONS – OVERALL

Language is inseparable from imagery and this is because of how it evolved.

It does have the dual reality described by Wundt, consisting of both an instantaneous (imagery) and a successive (linguistic-social) component. This ‘essential doubleness’ was also Saussure’s insight in his final notes on language. Without recognizing it as such, these great thinkers intuitively recaptured in part the evolutionary history of our species, as it has been affected by language.

According to this scenario, the dual cognitive mode of language was a product of how a capacity for language evolved, and how meaning came to control motor movements and their orchestration in the brain.