Growth points from the very beginning

David McNeill[1], Susan D. Duncan[1], Jonathan Cole[2], Shaun Gallagher[3] and Bennett Bertenthal[4]
[1]Department of Psychology, University of Chicago / [2]University of Bournemouth and Clinical Neurological Sciences, Clinical Neurophysiology, Poole Hospital, U. K. / [3]Department of Philosophy, University of Central Florida / [4]Department of Psychological and Brain Sciences, Indiana University

Early humans formed language units consisting of global and discrete dimensions of semiosis in dynamic opposition, or 'growth points.' At some point, gestures gained the power to orchestrate actions, manual and vocal, with significances other than those of the actions themselves, giving rise to cognition framed in dual terms. However, our proposal emphasizes natural selection of joint gesture-speech, not 'gesture-first' in language origin.

Keywords: social interaction, gesture, pantomime, growth point theory, thinking-for-speaking

Across time scales

Arbib and Bickerton, in their call for contributions to this special issue, posed the following puzzle: did protolanguage consist of units each having the semantic scope of a single word such as a noun or verb in present-day languages, or of holophrastic words, each with the synthetic scope of modern sentences?

We argue that the 'holophrasis *versus* compositionality' issue can be explored from an entirely novel perspective if co-verbal gesturing is fully taken into account. Our concern is with the evolution of the language system in which imagery and codified linguistic forms are "two aspects of utterance" (Kendon, 1980). Our claim is that this required a new mode of cognition. We claim that gesture (or, more broadly speaking, global-imagistic thinking) is a fundamental aspect of the human language system, and that an adequate account of human language evolution must posit as the end state of the evolutionary process a system that integrates

Interaction Studies 9:1 (2008), **117–132**. DOI 10.1075/is.9.1.09mcn
ISSN 1572–0373 / E-ISSN 1572–0381

compositional linguistic units and embodied global, synthetic, imagistic thinking. This in turn constrains possible evolutionary trajectories, which we explore.

Speech-synchronized gestures offer insights into the mental processes that modern-day language use engages. The nature of those processes suggests what may have enabled protolanguage. In modern humans, these gestures are integral components of language, not merely accompaniments. They are semantically and pragmatically co-expressive with speech, not redundant.[1] They are frequent — about 90% of spoken utterances in descriptive discourse are accompanied by them (Nobe, 2000) — and occur in similar form across speakers of many languages. We have observed such gestures in speakers of more than 20 cultures, including so-called 'high-gesture' cultures (e.g., Neapolitan). An inference to be drawn from observations of ubiquitous, universal, co-verbal gesturing is that thinking in terms of language utilizes two cognitive modes at once: analog imagery (visual, spatial, motoric) and discrete-categorial linguistic forms. In the following famous passage, Wundt a century ago expressed a similar perspective concerning the mental processes on which we focus:

> "From a psychological point of view, the sentence is both a simultaneous and a sequential structure. It is simultaneous because at each moment it is present in consciousness as a totality even though the individual subordinate elements may occasionally disappear from it. It is sequential because the configuration changes from moment to moment in its cognitive condition as individual constituents move into the focus of attention and out again one after another." (Wundt in Blumenthal, 1970)

Wundt here speaks of sentences but, as we explain below, synchronized, co-expressive gestures *and* speech constitute this duality, permitting us to observe it directly.

Gestures and speech — Two simultaneous modes of semiosis

Figure 1 illustrates synchronous co-expressive speech and gesture in narration.[2] (The speaker had just watched a cartoon and was recounting it to a listener from memory. We explained that the task was storytelling and did not mention gesture). The speaker was describing an event in which one character (Sylvester) attempted to reach another character (Tweety) by climbing up the inside of a drainpipe to a window where Tweety perched. The speaker said, "and he goes up thróugh the pipe this time." Co-expressively with "up" her hand rose; with "thróugh" her fingers spread outward to create an interior space. The upward movement and the opening of the hand were simultaneous and synchronized with "up thróugh," the linguistic package that carried the related meanings. The prosodic emphasis on

"thróugh," highlighting interiority, is matched by the added complexity of the gesture, the spreading and upturning of the fingers. What we mean by co-expressivity here is this joint highlighting of the ideas of rising and interiority, plus their joint contribution to "communicative dynamism" (Firbas, 1971).[3]

Figure 1. Gesture combining entity, upward movement and interiority in one symbol.

However, note also the differences between the two types of semiosis. Speech componentializes the event: a directed path ("up") plus the idea of interiority ("through"). This analytic segregation further requires that direction and interiority be concatenated, to obtain the composite meaning of the whole. In contrast, gesture is a synthesis. The whole emerges as one symbol. The semantic elements in speech are simultaneously aspects of this imagery whole. No concatenation is required. Meaning determination moves from whole to parts, not from parts to whole. The effect is a uniquely gestural way of packaging meaning — something like "rising hollowness." Thus, speech and gesture, co-expressive but non-redundant, represent one event (climbing up inside the pipe) in two forms: analytic/combinatoric and global/synthetic — at the same instant.

The kind of gesture we refer to is 'gesticulation.' This is gesture that is incomplete without speech accompaniment. Other types of gestures also may be part of communication but relate to speech in different ways. These differences are summarized in 'Kendon's Continuum' (first arrayed and named as such in McNeill, 1992; based on Kendon, 1988).

Kendon's continuum

Spontaneous Gesticulation → Language-slotted → Pantomime → Emblems → Signs

As one goes from gesticulation to sign language the relationship of gesture to speech changes:

- The obligatory presence of speech declines.
- Language-like properties increase.
- Socially regulated signs replace self-generated form-meaning pairs.

Language-slotted gestures have a different timing relationship from gesticulation with speech. For example in, "he goes [-]," a gesture synchronizes with a momentary pause in speech; a vacant grammatical slot. Here gesture substitutes for speech. An emblem is a culturally established morpheme (or semi-morpheme, because it does not usually have syntagmatic potential), such as the "OK" sign and others. Emblems can occur with or without speech. Pantomime is gesture without speech, often in sequences and usually comprised of simulated actions.[4] Sign languages are full, socially constituted, non-spoken languages. Even though 'gesticulation' (hereafter, 'gesture') is only one point on the Continuum, it dominates gesture output in storytelling, living space description, academic discourse (including prepared lectures) and conversation. Commonly 99% if not all gestures in such contexts count as 'gesticulation'.

Co-occurrence of gesticulation and speech involves simultaneous analog-imagistic and categorial-linguistic semiosis. The simultaneity of unlike semiotic modes reveals a 'double essence' of language — a term that Saussure, in notes discovered only recently, introduced to replace the *langue*/*parole* contrast (Saussure, 2002; see Harris, 2003). Although Saussure did not consider gesture, we propose that gesture is appropriate to his concept; the double essence is specifically *carried* by gesticulation plus linguistic encoding in cognition.[5]

Spontaneous gestures and encoded linguistic forms naturally contrast semiotically. Gesture is global and synthetic. Linguistic code is analytic and combinatoric. *Global* refers to the fact that the determination of meaning in a gesture proceeds top-down. The dimensions of meaning of the Figure 1 gesture (the hand means Sylvester, motion upward 'ascent', the fingers outspread 'interiority', etc.) are determined by the meaning of the whole: 'Sylvester as rising hollowness'. These dimensions are identifiable as parts only in the meaning landscape of the whole. They have no independent existence. It is not that gestures lack meaningful features but that features have no categorically-contrastive significance. Gesture features of the kind we are describing are in a globally-determining context.

This gestural mode contrasts with the bottom-up determination of meanings in sentences consisting of morphemes. In a synchronic sense, language must be conceived of as a system of such elemental morphemes, each having the potential to combine into larger wholes. This is so even if linguistic elements have ambiguities and/or graded qualities.

We propose that the first stages of language meshed global-'synthetic' gestural imagery with analytic-combinatoric, most likely vocal, signs. Even a single sign may have combinatoric potential if its significance seems incomplete. In terms of the holophrastic-combinatoric debate, therefore, we propose that both dimensions of semiosis were simultaneously present. *Synthetic* refers to the fact that a single gesticulation concentrates into one symbolic form distinct meanings that might be distributed over an entire construction ("he" + "goes" + "up" + "through" for example). Sign languages, of course, involve imagery but the imagery is either synchronically structured or integrated with synchronic structures (cf. Liddell, 2003). They are conventionalized so that even if their 'etymology' involves imagery, this imagery may not be used within the sign language.

The growth point

The GP is an irreducible, 'minimal unit'[6] of imagery-language code combination. It is the smallest packet of an idea unit encompassing the *unlike* semiotic modes of imagery and linguistic encoding. The GP carries the Saussurian double essence in the domain of psycholinguistic processing. A GP is empirically recoverable, inferred from speech-gesture synchrony and co-expressiveness.[7] The temporal and semantic synchronies represented in Figure 1 imply a GP built on the idea of rising interiority. We infer the simultaneous presence of the idea of ascent inside the pipe in two unlike semiotic modes. Even when the information ('semantic content') in speech and gesture is similar, it is formed according to contrasting semiotic modes. Simultaneous unlike modes create instability. Instability fuels thinking-for-speaking as it seeks resolution (McNeill & Duncan, 2000).[8]

The GP is so named because it is a distillation of a growth process — an ontogenetic-like process but vastly sped up and made functional in online thinking-for-speaking. According to this framework, it is the initial unit of thinking-for-speaking (Slobin, 1987) out of which a dynamic process of utterance-level and discourse-level organization emerges. Imagery and spoken form are mutually influencing. It is not that imagery is the input to spoken form or spoken form is the input to imagery. The GP is fundamentally both.[9]

For modern humans, stability comes from 'unpacking' the growth point into grammatical structures (or viable approximations thereto). A surface linguistic form emerges to embody its GP in this maximally stable form. This role of grammar — unpacking and supplying 'stop-orders' for the changes initiated by imagery-linguistic code instability — is an important clue about protolanguage. Instability would have been present in this situation at the time of the earliest linguistic explorations too. Imagery-linguistic encoding creates pressure for grammar to stabilize the process.[10] Unstable semiotic embodiments could coexist with

other factors, such as the need to limit ambiguity, and together provide pressure for the development of grammatical constructions (cf. Arbib 2005). In Figure 1, "up through" is analytic: up-ness and interiority are separated. The words also have syntagmatic values acquired from combinations within and beyond the phrase. The gestural image embodies this information — 'Sylvester as rising hollowness' — without any combinatoric value. Unpacking resolves the tension of semiotic modes. The utterance, "(he) goes up thróugh it" accommodates both the linguistic encoding and the imagery.

A final point is that we can fully understand what motivates any image-speech combination only with reference to how a GP relates to its context of occurrence. The GP-to-context relationship is mutually constitutive. The GP is a point of differentiation from the context. The speaker represents the context to make this differentiation possible within it. A robust phenomenon concerning gesture is that the form and timing of gesture select just those features that differentiate the psychological predicate in a context that is at least partly the speaker's creation (see McNeill, 2005, pp. 108–112). The 'double essence' of language includes incorporation of context.

Next we offer evidence for a brain link that evolved specifically to sustain the duality of holistic imagery and discrete linguistic encoding.

A thought-language-hand brain link

The IW case

An implication of the GP hypothesis is that imagistic thinking, of the kind materialized in gestures, is an integral component of language production. This hypothesis would be supported by a speaker whose physical condition would otherwise prevent instrumental actions from occurring normally, yet still gestures with speech. 'IW' is such a speaker. Due to an autoimmune-induced, large sensory fiber neuronopathy at age 19, IW was deafferented over his entire body below the neck. With great effort, IW, now in middle age, has reestablished control of his motor system using cognition and vision in the complete absence of proprioception and spatial position sense (see Cole, 1995). If his vision of his own actions is occluded, IW cannot perform instrumental actions. He can, however, without any other sense of what his hands are doing, perform morphokinetically well-formed gestures that synchronize with speech as normal. This dissociation of instrumental action and gesture reveals a thought-language-hand link in the human brain not otherwise discernible. In instrumental action, orientation to specific objects in the world directs action. For gesture in a language use context, however, it is thought

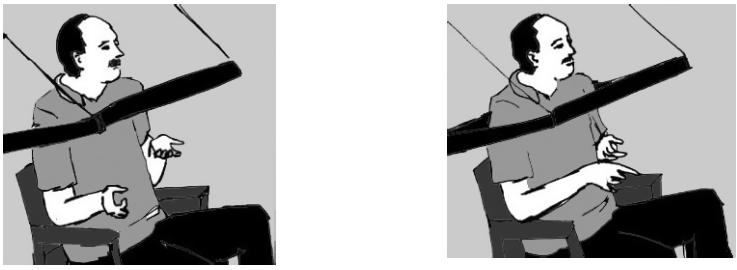

Figure 2. IW coordinated two-handed iconic gesture without vision.

that organizes the same motor system. An implication is that part of language evolution involved reorganization of the brain's motor system to take thoughts and ideas rather than representations of real-world objects as input to motoric activity.

Figure 2 shows synchronized speech-gesture co-expressiveness in IW's spontaneous language use. He created a coordinated two-handed gesture accompanying his spoken narrative while from the shoulders down he was under a blind[11] that concealed his hands while permitting movement.

His left hand is Sylvester and his right hand is a trolley in pursuit. IW was saying, "and the tram **caught** him up." His right hand moved to the left (boldface) in synchrony with the co-expressive "caught". Moreover, a poststroke hold (underlining) continued the image through "him" and "up", maintaining full synchrony with the co-expressive speech. Keep in mind that synchrony and co-expressivity were achieved without vision or proprioceptive or spatial feedback. Visuo-spatio-motoric imagery alone drove this coordinated use of his hands.

Another indication of a thought-language-hand link is that IW, without vision, can modulate the speed at which he presents meanings in speech and gesture *in tandem*. As IW's speech slows down, his gesture slows down equally. Still with the blind during a conversation with co-author Jonathan Cole, IW slowed his speech at one point by a third (paralinguistic slowing) but speech and gesture remained in synchrony (for more detail, see McNeill, 2005, pp. 243–244). The gesture was the two hands repetitively rotating outward and inward, apparently a metaphor for the idea of a process. This metaphoric significance is consistent with the synchronized speech. The hands rotated only while IW was saying, "I'm starting to get into" and "I'm starting to use." He paused between the first (normal speed) and second (reduced speed) rotations as he said, "and that's because," indicating that the rotation and any associated phonetic linkages were specifically organized around the metaphor.

Speech and gesture, slowing together, could mean that the realization of a GP entails slowing the potentially faster unit to the duration of the slower one, a principle also invoked in models of reach-grasp coordination. If the hands move out-

ward in synchrony with a breath pulse accent, for example, the association could be maintained over a range of speeds. Such a pacesetter accordingly could be activated by the thought-language-hand link and co-opted by a significance other than the action of rotation itself. Again, recall that IW had no idea where his hands were. An explanation for his performance is that gestures and speech were combined online and controlled by meanings other than those organized by the attainment of real-world ends; the process metaphor. In terms of the GP hypothesis, any change in the length of time a GP is active affects the linguistic and gestured components of IW's utterances in tandem.

Along the same line, when IW is told to imitate actions or other gestures, he spontaneously begins to speak and if he is told not to speak, the imitation is impaired. What advantage does IW's gesture have for him? Is it to make him appear normal, since it was initially absent after the deafferentation (Cole, 1995), or is it because elaboration of the gesture is necessary for the thought/language system? If the latter, the advantage is in the central motor/representation systems, since IW has no feedback of gesture. It is an internal feed forward advantage.

The IW case implies that the "know-how" of gesture is not the same as that of instrumental action. To understand this implication, sit facing a table and put your hands out of sight under its surface. Open and close one hand, extend one finger, then reach over to the other hand and touch it. You know at all times where your hands are and what they are doing but IW would not. For him the required spatial and proprioceptive information are absent. Yet he performs co-expressive gestures that synchronize precisely with speech under these very conditions. It is not correct to say that IW gestures normally as a result of having already developed the 'skill' by age 19, the time of his illness. He had those same years to develop his capability with instrumental actions, yet now falters when performing those without vision, even at the level of morphokinetic accuracy. What the hypothesized GP-type cognition explains is how, currently, he synchronizes co-expressive speech and gesture without visual guidance, exactly as normal. In this mode of cognition, speech and gesture are effectively one.

Next we suggest a mechanism by which such a brain link evolved specifically to sustain GPs.

GPs and language evolution

We suggest that the initial protolanguage would have comprised: (a) recurrent, 'code'-based forms (potentially listable and portable from speech situation to situation) and (b) context-determined, global, holistic imagery-based forms (neither listable nor portable). This combination would have, we propose, emerged in GP-

like units of cognition. The code component could be simple, even single symbols, but with analytic and combinatoric potentials like those mentioned. It is a new cognitive capacity that we see evolving.

The opposition of semiotic modes carrying the double essence demands that the linguistic side be socially constituted: shared, discrete, repeatable, combinable and listable. From the social direction come the semiotic properties that are unlike imagery. Imagery in turn has its source in context-specific thought processes. Rapidly merging unlike semiotic modes — imagery and *langue*-like encodings — during ongoing speech is the trick we evolved as a species. It is a necessary foundation for human language. So at the origin, a breakthrough was to create a social standard of predictable symbols that could combine with individually constituted, contextually situated and ephemeral imagery. Our focus now is on the immediate steps that might have taken place in the origin of language; not the rise of linguistic systems as such, but the brain mechanisms creating a thought-language-hand link. What would such more immediate steps have been? Our proposal is that this link was a new way to organize sequences of movements in Broca's area. The crucial new step was the co-opting of these areas by significances other than those of actions themselves. And how did other significances gain this power? Along with Arbib (2005) and McNeill (2005) we propose that *it was by making mirror neuron circuits respond to one's own gestures*.

'Mead's Loop' and mirror neurons

George Herbert Mead wrote that, "[g]estures become significant symbols when they implicitly arouse in an individual making them the same response which they explicitly arouse in other individuals" (1974, p. 47). Thus, gesture implies a 'social other', real or virtual. What was selected, in this view, is a capacity, not present in other primate brains, for mirror neurons to respond to *one's own gestures* as if they were social objects (cf. Cohen, 1977, who observed significantly less gesture when people speak into a tape recorder, compared to talking on the phone). Imagery in the form of gesture can be shared (Kimbara, 2006). Also, imagery that implies a social other is oriented to the socially-constituted semiotic of language, meshing smoothly with *langue*-like encodings in the duality we have described — global/synthetic imagery combined with analytic/combinatoric *langue*. Mead's Loop creates a new basis for organizing actions. At the motor level, the Loop provides a way for significant imagery to enter and be available to orchestrate Broca's area, giving the gesture the property of 'chunking': a chunk of linguistic output organized around significant imagery rather than an instrumental action. We hypothesize that, evolutionarily, Mead's Loop co-opted the brain's mirror neuron circuit.

This 'Mead's Loop' explains how gestures could reorganize the part of the brain in which complex actions are orchestrated — ventral premotor cortex and inferior frontal gyrus in the modern brain — so that significances other than actions themselves organize movements; e.g., rotation that abstractly represents a process, rather than the same motor neurons and brain areas signaling the hands to rotate an actual object as a goal-directed action. It is likely the same neurons are activated, but we are suggesting that the inputs to them differ. This is of course Broca's area, the repository of motor mirror neurons — neural circuits not only implicated in the execution of one's own movements but also capable of recognizing the goal-directed actions of others (Rizzolatti & Arbib, 1998; Arbib, 2005).[12] Specifically, what we infer about the thought-language-hand link in the IW case is that the input to the mirror neurons is not sensory information from tangible objects; rather, some form of imagery from higher-level cognition. This imagery differs from the physical objects of goal-directed actions because shared attention is not sufficient to ground the meaning. In the case of a visible object of goal-directed action, one can observe another person reach for the object and know the intended goal via activation of the mirror system. By contrast, in the case of communicating a thought, one observes another gesture, but the goal is only implicit. It is therefore important that Mead's Loop first be established to allow for a shared meaning.

We submit that this was one step in the emergence of language. Mead's Loop specifically explains how the thought-language-hand link revealed by the IW case could have evolved. It is precisely this link that, according to the logic of Mead's Loop, natural selection would promote.

But NOT 'gesture-first'

The Mead's Loop mechanism, however, does not mesh well with the claim (e.g., Armstrong et al., 1995, Corballis, 2002, Arbib, 2005; many others, informally) that language began as gesture, a recently reactivated, 18th Century theory (Condillac; see Harris & Taylor, 1989). 'Gesture-first' posits that early humans first developed something like a sign language. Then as we, and our linguistic conceptualizations, became more complex, speech supplanted the gesture-based system, with gesture 'scaffolding' the transition (Arbib, 2005). Such accounts, however, are founded on an inadequate analysis of the current state of the language evolution process.

Gesturing is integral to modern-day language and is evidence of the distinctive mode of cognition that we have described here. 'Gesture-first' provides instead sign systems (Armstrong, et al. 1995) or pantomime (Arbib, 2005) as 'stepping-stones.'

An assumption (unspoken) of such theories is that gesture (pantomime) would have provided an 'easy entrée' to protolanguage because the modality af-

fords iconic depiction. There may indeed have been pantomimes without vocalizations for communication at the dawn, in which case pantomime could have had its own evolution, landing at a different point on Kendon's Continuum, reflected today in the different temporal relationship to speech: alternating rather than synchronous. With gesticulation the individual speaker constructs a combination of speech and gesture, combined at the point of maximal co-expressiveness. In pantomime none of this occurs. As a 'stepping stone,' it could not have led to such combinations. There is no co-construction with speech, no co-expressiveness, and timing is different, if there is speech at all. The very same movement — that in Figure 1, for example — may occur as a pantomime or as a gesticulation. Whether or not the speaker creates the cognitive unit with which to combine the movement with speech is the key discriminating factor.

'Scaffolding', if it occurred, would seem to entail that pantomime/sign and speech at some point crossed paths. There is a model of this co-existence for us to examine. Emmorey et al. (2005) observe frequent pairings of signs and speech by hearing ASL/English bilinguals. While 94% of such pairings are signs and words translating each other, 6% are not mutual translations. In the latter, sign and speech collaborate to form sentences, half in speech, half in sign. For example, a bilingual says, "all of a sudden [LOOKS-AT-ME]" (from a Sylvester and Tweety cartoon narration; capitals signify signs simultaneous with speech). This could be 'scaffolding' but notice that it does not create the combination of unlike semiosis that we have described. Signs and words are of the same semiotic type — segmented, analytic, repeatable, listable, and so on. There is no global-synthetic component, and no built-in merging of analytic/combinatoric forms and global synthesis. Of course, ASL/English bilinguals have the ability to form GP-style cognitive units. But if we imagine a transitional species evolving this ability, the Emmorey et al. model suggests that scaffolding did not lead to GP-style cognition; on the contrary, it implies two analytic/combinatoric codes dividing the work. If we surmise that an old pantomime/sign system did scaffold speech and then withered away, this leaves us unable to explain how gesticulation, with the special cognitive process we have described, emerged and became engaged with speech. We conclude that scaffolding, even if it occurred, would not have led to current-day speech-gesticulation linkages.

Our claim, in other words, is that Kendon's "two aspects of utterance" requires the qualitatively different mode of cognition we observe when speech and gesticulation combine, and that gesture-first, in any of its versions, cannot explain the origin of this cognition. Moreover, we observe in the gestures of modern humans (see McNeill 1992, 2005; McNeill & Duncan, 2000) a sensitivity to discourse content and highly selective expression of discourse focal elements that undercuts an

'easy-entrée' assumption. The discourse-contextualized nature of gestures is compatible with the GP formulation; in fact, is required by GPs for differentiation.

We are advocating the view that language evolved when the capacity to form speech-gesture units was naturally selected. Speech and gesture would have evolved together (cf. Volterra, et al., 2005). The plausibility of this hypothesis is bolstered by the observation that chimpanzees show hand dominance for gestures only when the movements co-occur with vocalization (Hopkins & Cantero, 2003). Barring independent evolution by chimps, such combinations would have existed in the last common human-chimp ancestor[13] and would have provided the raw material for co-opting the motor area by imagery, thought and language.

Pollick & de Waal (2007) report that chimps and bonobos, " … use brachiomanual gestures more flexibly across contexts than they do facial expressions and vocalizations" (p. 8187; also Tomasello & Call, 1997). Although they regard this difference as support for gesture-first, it is equally compatible with what we have argued is the more plausible hypothesis of gesture-speech evolving jointly, not sequentially. Gestures sensitive to context are just what Mead's Loop 'wants'. If gesture and vocalization were linked pre-adaptively, as above, the impact of Mead's Loop on Broca's area could orchestrate speech and gesture jointly. What bonobos and chimps may lack is precisely evolution via Mead's Loop; a reason, perhaps, they have not advanced beyond gesture.

To sum up, gesture-first predicts what did not evolve (pantomime as a step toward language, gesture a vestige to be shown the door)[14] and does not predict what did evolve (the instability of simultaneous semiotic modes, the 'double essence', gesticulation embodying discourse organization). In the Mead's Loop model, in contrast, speaking could not have evolved without simultaneous gesture, and gesture could not have evolved without its duet with speech.

Conclusions

Framing the evolution of language question as one of when GP-type cognition evolved skirts the holophrasis versus combinatorics puzzle. An unstable meshing of unlike semiotic modes craves stability. Instability arises even when primitive linguistic encodings mesh with imagery. From this vantage point, we claim that sentences continue the evolution that started with GPs. 'Social-fact' encodings arise in the act of sharing information, creating a 'discreteness filter' such that the semiotic properties of segmentation and potential for combination arise automatically (Freyd, 1983). In GPs, such encodings (initially simple) already interlock with imagery. Sentences, whatever their complexity, stabilize GPs by adding information. Evolutionarily, this step required two cultural and/or biological extensions

of the impulse to stabilize semiotic opposites: *constructions*, which differentiate holistic meanings by adding semantic and syntactic frames (Goldberg, 1995) and *recursive embeddings*, which elaborate meanings by combining constructions (cf. Hauser, et al., 2002). Thus, holophrasis and composition would have arisen from GPs at the dawn and made the development of grammar advantageous.

We hold that language (neither speech nor gesture preceding) evolved as part of a cognitive mode integrating holistic imagery and discrete code. This capability ushered in new modes of action (cf. Vygotsky, 1987), which we propose were initially speech and gesture, in which orientation to social interaction is inherent, conferring adaptive advantages, and so was naturally selected. Some of these steps may have required biological adaptations. We would expect these adaptations to occur in the system of motor control — ever more complex ways of orchestrating movements (oral, laryngeal, manual), under significances in which orientation to social interaction is inherent — taking values other than those of the actions themselves (chewing, screeching, manipulating, etc.). Via gestures, significant imagery orchestrates motor behavior, and this was a key step in the origin of protolanguage.

Notes

1. Our discussion does not concern signs, salutes, or "emblems" (e.g., Ekman & Friesen, 1969).

2. More extensive accounts are in McNeill (1992) and McNeill (2005).

3. Computer art from video by Fey Parrill, Ph.D.

4. What distinguishes pantomime from gesticulation is that the latter, but not the former, is integrated with speech. Gesticulation is a dimension of speaking. Pantomime, if it relates to speaking at all, does so as a 'gap filler'. Speech-gesticulation combinations are cognitive constructions, and occur where speech and gesture are co-expressive of the same idea. Movement by itself offers no clue to whether a gesture is 'gesticulation' or 'pantomime'; what matters is whether the two modes of semiosis simultaneously co-express one idea unit.

5. Harris (2003) emphasizes that it is *langage*, not *langue*, that Saussure intended — the full dynamic semiological phenomenon.

6. The concept of a 'minimal unit' with the property of being a whole is from Vygotsky (1987, pp. 4–5).

7. A growth point is inferred (not 'operationally defined') from a) gesture form, b) coincident linguistic segment(s), c) co-expression of the same idea unit, and d) what Vygotsky (1987, p. 243) termed a 'psychological predicate' in the immediate context of speaking (of which, more below).

8. The reasons why semiotic opposition creates instability and initiates change include:

a. *conflict* (between semiotic modes: analog imagery/analytic categorical), and
b. *resolution* (through change: fueling thinking-for-speaking, seeking stability).

Simultaneous semiotic modes comprise an inherently dynamic psycholinguistic model.

9. Comparison of GPs to other hypotheses given in McNeill (2005, Chapter 4.3).

10. When gesture and speech synchronize, as in Figure 1, the two modes are in direct contact. If there is less than perfect synchrony, the 'double essence' can still urge unpacking. The ultimate criterion is whether an idea is embodied in two forms (with or without different aspects of the idea) and this creates instability.

11. Nobuhiro Furuyama suggested the blind. The blind itself was designed and built by David Klein.

12. Bertenthal et al. (2006) showed that the human mirror system is sensitive to both intransitive actions and gestures.

13. Or further back: Fogassi & Ferrari (2004) have identified neural mechanisms in monkeys for associating gestures and meaningful sounds, which they suggest could be a pre-adaptation for articulated speech.

14. Rizzolatti and Arbib (1998), for example, write, "Manual gestures progressively lost their importance, whereas, by contrast, vocalization acquired autonomy, until the relation between gestural and vocal commuication inverted and gesture became purely an accessory factor to sound communication" (p. 193).

References

Arbib, M. A. (2005). From monkey-like action recognition to human language: An evolutionary framework for neurolinguistics. *Behavioral and Brain Sciences, 28*,105–124.

Armstrong, David F., Stokoe, William C., & Wilcox, Sherman E. (1995). *Gesture and the nature of language*. Cambridge: Cambridge University Press.

Bertenthal, Bennett, Longo, Matthew R. & Kosobud, Adam. (2006). Imitative response tendencies following observation of intransitive actions. *Journal of Experimental Psychology: Human Perception and Performance*, *32*, 210–225.

Blumenthal, Arthur (Ed. and Trans.). (1970). *Language and psychology: Historical aspects of psycholinguistics*. New York: John Wiley & Sons Ltd.

Cohen, Akiba A. (1977). The communicative function of hand illustrators. *Journal of Communication, 27*, 54–63.

Cole, Jonathan. (1995). *Pride and a daily marathon*. Cambridge, MA: MIT Press.

Corballis, Michael C. (2002). *From hand to mouth: The origins of language*. Princeton, NJ: Princeton University Press.

Ekman, Paul, & Friesen, Wallace (1969). The repertoire of non-verbal behavior: Categories, origins, usage and coding. *Semiotica* 1. 49–98.

Emmorey, Karen, Borinstein, Helsa B. & Thompson, Robin (2005). Bimodal bilingualism: Code-blending between spoken English and American Sign Language. In James Cohen,

Kara T. McAlister, Kellie Rolstad & Jeff MacSwan (eds.), *Proceedings of the 4th International Symposium on Bilingualism* (pp. 663–673). Somerville, MA: Cascadilla Press.

Firbas, Jan. (1971). On the concept of communicative dynamism in the theory of functional sentence perspective. *Philologica Pragensia, 8*, 135–144.

Fogassi, Leonardo & Ferrari, Pier Francesco. (2004). Mirror neurons, gestures and language evolution. *Interaction Studies, 5*, 345–363.

Freyd, Jennifer J. (1983). Shareability: The social psychology of epistemology. *Cognitive Science* 7: 191–210.

Goldberg, Adele. (1995). *Constructions: A construction approach to argument structure*. Chicago: University of Chicago Press.

Harris, Roy. 2003. *Saussure and His Interpreters*, 2nd ed. Edinburgh: Edinburgh University Press.

Harris, Roy & Taylor, Talbot J. (1989). *Landmarks in linguistic thought: the western tradition from Socrates to Saussure*. New York: Routledge.

Hauser, Marc, Chomsky, Noam, & Fitch, W. Tecumseh (2002). The language faculty: What is it, who has it, and how did it evolve? *Science, 298*, 1569–1579.

Hopkins, William D. & Cantero, Monica. (2003). From hand to mouth in the evolution of language: the influence of vocal behavior on lateralized hand use in manual gestures by chimpanzees (*Pan troglodytes*). *Developmental Science*, 6, 55–61.

Kendon, Adam (1980). Gesticulation and speech: two aspects of the process of utterance. In M. R. Key (ed), *The Relationship of Verbal and Nonverbal Communication* (pp. 207–227). The Hague: Mouton and Co.

Kendon, Adam (1988). How gestures can become like words. In F. Poyatos (ed.), *Cross-cultural perspectives in nonverbal communication* (pp. 131–141). Toronto: Hogrefe.

Kimbara, Irene. (2006). *Interpersonal influences on gesture production : evidence for gesture form convergence across speakers in dyadic interaction*. Unpublished doctoral dissertation, University of Chicago.

Liddell, Scott (2003). *Grammar, Gesture and Meaning in American Sign Language*. Cambridge: Cambridge University Press.

McNeill, David. (1992). *Hand and mind: What gestures reveal about thought*. Chicago: University of Chicago Press.

McNeill, David. (2005). *Gesture and thought*. Chicago: University of Chicago Press.

McNeill, David, & Duncan, Susan D. (2000). Growth points in thinking for speaking. In D. McNeill (ed.), *Language and gesture* (pp. 141–161). Cambridge: Cambridge University Press.

Mead, George Herbert. (1974). *Mind, self, and society from the standpoint of a social behaviorist (C. W. Morris ed. and introduction). Chicago: University of Chicago Press.*

Nobe, Shuichi. *(2000).* Where do most spontaneous representational gestures actually occur with respect to speech? In D. McNeill (ed.), *Language and gesture* (pp. 186–198). Cambridge: Cambridge University Press.

Pollick, Amy S. & de Waal, Frans B.M. (2007). Ape gestures and language evolution. *Proceedings of the National Academy of Sciences, 104*, 8184–8189.

Rizzolatti, Giacomo & Arbib, Michael (1998). Language within our grasp. *Trends in Neurosciences, 21*, 188–194.

Saussure, Ferdinand de. 2002. *Écrits de linguistique général* (compiled and edited by S. Bouquet and R. Engler). Paris: Gallimard.

Slobin, Dan I. (1987). Thinking for speaking. In J. Aske, N. Beery, L. Michaelis, & H. Filip (eds.), *Proceedings of the Thirteenth Annual Meeting of the Berkeley Linguistic Society* (pp. 435–445). Berkeley: Berkeley Linguistic Society.

Tomasello, Michael & Call, Josep (1997). *Primate cognition*. New York: Oxford University Press.

Volterra, Virgina, Caselli, Mari Cristina, Caprici, Olga, and Pizzuto, Elena. (2005). Gesture and the emergence and development of language. In M. Tomasello & D. Slobin (eds.), *Beyond Nature-Nurture: Essays in Honor of Elizabeth Bates*, (pp. 3–40). Mahway, N.J.: Erlbaum.

Vygotsky, Lev S. (1987). *Thought and language*. Edited and translated by E. Hanfmann and G. Vakar (revised and edited by A. Kozulin). Cambridge: MIT Press.

Authors' addresses

David Mc Neill
Department of Psychology
University of Chicago
5848 S. University Ave
Chicago, IL60637

Susan D. Duncan
Department of Psychology
University of Chicago
5848 S. University Ave
Chicago, IL60637

Jonathan Cole
Poole Hospital
Longfleet Road
Poole BH15 2JB
United Kingdom

Shawn Gallagher
Department of Philosophy and the Cognitive Science Program
University of Central Florida,
Orlando
FL32816-1352

Bennett J. Bertenthall
Department of Psychological and Brain Sciences
Indiana University
Bloomington
IN 47405-7000

About the authors

David McNeill's 1992 book, *Hand and Mind*, received the Laing Prize in 1994 from the University of Chicago Press.

Susan D. Duncan is a pycholinguist and has published widely on the topic of gesture.

Jonathan Cole is a clinical neurophysiologist and has published books by the MIT Press on living with sensory loss, facial difference and quadriplegia.

Shaun Gallagher is a philosopher and cognitive scientist. His most recent book is *How the Body Shapes the Mind* (2005).

Bennett I. Bertenthal has widely published on the development of perception and action, developmental cognitive neuroscience, multimodal communication, cyberinfrastructure, and science policy.