Gesture and Thought

David MCNEILL
University of Chicago

Abstract. Both a synopsis and extension of *Gesture and Thought* (the book), the present essay explores how gestures and language work together in a dialectic. In this analysis the 'purpose' of gesture is to fuel and propel thought and speech. A case study illustrates the dependence of verbal thought on context and how it functions. Problems for computational modeling, the presence and absence of gesture 'morphemes, and speculation on how an imagery-language dialectic evolved are also provided.

Keywords. Gesture, growth points, imagery-language dialectic, dynamic dimension of language, psychological predication

1. Dialectic

Gesture and Thought, a book of the same title as this essay [1], presents a new conception of language: language as an imagery-language dialectic in which the role of gestures is to provide imagery for the dialectic. Gesture is an integral component of language in this conception, not merely an accompaniment or ornament. Such gestures are synchronous and co-expressive with speech, not redundant, and not signs, salutes, or emblems. They are frequent—about 90% of spoken utterances in descriptive discourse are accompanied by them [2]. They synchronize with speech at the point where speech and gesture co-expressively embody a single underlying meaning, a meaning that is the point of highest communicative dynamism at the moment of speaking. A host of phenomena testify to a tight bond, to the point of fusion, of the speech-gesture combination.[1] The synchrony of speech forms and gestures creates the conditions for an imagery-language dialectic. A dialectic implies:

- A conflict or opposition of some kind, and
- Resolution of the conflict through further change or development.

The synchronous presence of unlike modes of cognition, imagery and language, co-expressive of the same underlying thought unit, sets up an unstable confrontation of opposites. It is this very instability that fuels thinking for speaking as it seeks resolution. Instability is an essential feature of the dialectic,

[1] Among them: 1) The disruption of speech flow caused by delayed auditory feedback does not interfere with speech-gesture synchrony: the cross-modal unit remains intact ([3], first DAF experiment). 2) The onset of a gesture stroke inoculates against clinical stuttering. The onset of stuttering, once a stroke has begun, causes immediate cessation of the stroke [4]. In both cases, stuttering and gesture stroke are incompatible. 3) Gestures and speech spontaneously exchange semantic complexity in memory – information presented in gesture may be recalled in speech but not in gesture [5] and information in speech recalled in gesture but not in speech [6]. 4) Congenitally blind speakers perform gestures even to a known blind listener [7]. That is, so strong is the speech-gesture bond, that speakers with no experience of gesture, speaking to listeners known to have no perception of gesture, perform gestures (presumably unwittingly) with the flow of speech.

and is a key to the dynamic dimension. The concept of an imagery-language dialectic extends a concept initiated by Vygotsky, in the 1930s [8].

This new conception also recaptures an insight lost for almost a century, that language requires two simultaneous modes of thought—what Saussure, in recently discovered notes composed around 1910 [9], termed the 'double essence' of language (although he expressed this without reference to gestures). Wundt [10], writing about the same time, had a similar insight in this famous passage:

> "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" (p. 21).[1]

Gesture and Thought focuses on the real-time actualization of thought and language, regarding language multimodally and in context—its dynamic dimension. On the dynamic dimension, language appears to be a process, not an object. On the crosscutting static dimension, it looks to be an object but not a process. In fact, both dimensions must be considered, as both are indispensable to a full theoretical explication of utterances. An important question is how they combine in real-time utterances.

2. Imagery

'Imagery', as intended here, is a symbolic carrier that *lacks duality of patterning* – to use Charles Hockett's term for one of the design features of language [11]; imagery lacks this, it is a symbolic form determined by meaning, not by a system of form contrasts or standards of good form. Imagery is actional as well as visuospatial. It is also non-photographic, since the form of the image is driven by meaning, not by external stimulation (or not only this).

2.1. And metaphor

Thanks to metaphoricity, imagery is not restricted to concrete references. Metaphoricity in gesture is a fundamental property [12]; it is not mere ornamentation. It expands imagery to encompass abstract meaning in an dialectic with linguistic form. A famous example is the 'conduit' metaphor, which appears in such purely verbal uses as "there was a lot in that book"; the image being that meaning is a substance and the book is a container [13, 14]. A conduit metaphor in gesture is the Palm Up Open Hand (PUOH), described by Cornelia Müller [15], where the palm 'holds' or 'contains' some discursive 'substance'. Via PUOH, totally abstract content can as imagery dialectically oppose co-expressive linguistic material. An example from a cartoon narration is a speaker saying "and the next scene is" – abstract content – and at the same time making a PUOH (the metaphor in gesture only, a frequent asymmetry).

[1] I am grateful to Zenzi Griffin for alerting me to this passage.

2.2. But not a morphology

But is the recurring gesture imagery in the PUOH the beginning of a gesture morphology? Some gestures, such as the Neapolitan ones described by Kendon [16], seem clearly structured as morphologies, but what of metaphors like PUOH? As explained, the imagery component of a dialectic lacks duality of patterning, whereas a morphology, in its usual definition, is exactly the sort of thing to which duality of patterning applies. The puzzle arises because, in the case of PUOH and other variants of the conduit, there *is* regulation of form – by the metaphor itself; there is a container or surface, into which 'substance' goes. Is this form requirement morphemic?

The key question is whether the form of the gesture arises only from significance (thinking of the structure of the story and using the conduit metaphor, I construe the next scene as a container) or is also structured on the level of form, qua form. In the PUOH case it appears that the metaphor is all that is required. Occam's razor applies: there is no warrant for a further hypothesis of form regulation beyond the metaphor.

Further evidence against a morphology of gesture in these situations is that when gestures recur in cartoon narrations there is no ritualization or streamlining. This is true not only of PUOH but of all sorts of iconic depictions of story characters and other situations. Gesture recurrences take place because the same imagery arises in much the same way each time. Streamlining however requires more. It demands the linking of the recurrences into a system of some sort. Absent this, recurrences are just imagery in more or less the same form time after time.

It is possible nonetheless to induce something like a true gesture morphology by breaking apart the gesture-speech combination. By outlawing speech, getting speakers to recount a tale without words, using gestures alone, standardized forms of gesture emerge spontaneously. For example, in a non-verbal rendition of the Snow White story (Ralph Bloom study [17], for description see [3]) the King and Queen signs had stable forms, and the contrast between them appeared on their first appearance. The King sign was initially made with a two-handed jagged encirclement of the head (the crown), followed directly by a bracing of arms at the side (muscles or flat-chested); the Queen sign was made with the same crown but with the two hands cupped upward for breasts. The gestures therefore immediately contrasted. They underwent extensive streamlining – the crown ultimately became a flick of a single hand oriented toward (but not necessarily at) the head, the breasts or muscles different orientations of one or both hands (palm toward chest for King, palm up for Queen). The crown and flat-chested vs. breasts features never disappeared during some 70 recurrences. Ritualization thus took place and preserved all the significant form contrasts. In other words, absence of speech was compensated for by a gesture morphology. However, with speech present there is no such pressure and system of form contrasts for gesture does not arise; instead, there is an instantaneous combination of unlike semiotic modalities.

3. A specific gesture type

The semiotic combinations are summarized in what I once termed Kendon's Continuum [3], named after [18]:

> Spontaneous Gesticulation (Mode 1) → Language-slotted (Mode 2) → Pantomime (Mode 2) → Emblems (Mode 2) → Signs

Mode 1=unwitting gestures, Mode 2=gestures intended as symbols (due to S. Duncan).

As one goes from gesticulation to sign language:

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

Gesticulation is the type of gesture analyzed in depth here; language-slotted gestures are also gesticulations but replace speech rather than synchronize with it ("he goes [gesture]" – the gesture timed to coincide with a vacant grammatical slot); pantomime is dumb-show and occurs without speech at all; an emblem is a culturally established morpheme (or semi-morpheme, because it does not usually have combinatoric potential) such as the "OK" sign and others, and occurs with or without speech; and sign languages are socially-constituted languages and do not combine with speech (American Sign Language and others).

Even though 'gesticulation' (hereafter, 'gesture') is only one point on the Continuum, in storytelling, living space descriptions, academic discourse (including prepared lectures) and conversations the overwhelming gesture type is gesticulation – commonly 99% if not totally (the propensity to adopt conventionalized emblems or 'quotable' gestures [16] varies across cultures; in the genres listed, among North American speakers at least, they are overwhelmingly absent). As the Mode 1/Mode 2 distinction indicates, gesticulations alone are unwitting, not intended as symbols. They are integrated with linguistic content into growth points and appear, to the speaker, to be an unbroken package of semiosis with it.[3]

4. The growth point

The smallest unit of the imagery-language dialectic is posited to be a 'growth point,' so named because it is theoretically the initial unit of thinking for speaking out of which a dynamic process of organization emerges. A growth point combines imagery and linguistic categorial content, and the theory is that such a combination initiates cognitive events. A growth point is an empirically recoverable idea unit, inferred from speech-gesture synchrony and co-expressiveness.

[3] The Continuum was elaborated into four Continua in [19].

4.1. A case study

An example recorded in an experiment (offered in part because of its ordinariness) is a description by one speaker of a classic Tweety and Sylvester escapade, which went in part as follows: "and Tweety Bird runs and gets a bowling ba[ll and drops **it do**wn the drainpipe]." Speech was accompanied by a gesture in which the two hands thrust downward at chest level, the palms curved and angled inward and downward, as if curved over the top of a large spherical object. At the left bracket, the hands started to move up from the speaker's lap to prepare for the downward thrust. Then the hands, at the very end of "drops," held briefly in the curved palm-down position, frozen in midair (the first underlining). Next was the gesture stroke—the downward thrust itself—timed exactly with "it down" (boldface). Movement proper ceased in the middle of "down," the hands again freezing in midair until the word was finished (the second underlining). Finally, the hands returned to rest (end of second underlining up to the right bracket). The two holds reveal that the downward thrust was targeted precisely at the "it down" fragment: the downward thrust was withheld until the speech fragment could begin and was maintained, despite a lack of movement, until the fragment was completed. Significantly, even though the gesture depicted downward thrusting, the stroke bypassed the very verb that describes this motion, "drops," the preparation continuing right through it and holding at the end—an explanation for this seeming overshoot is provided later.

The growth point was thus the fragment, "it down," plus the image of a downward thrust. Both sides of the growth point are essential, and are opposed dialectically in that the linguistic components have combinatoric potential and categorize the image; the imagery component embodies these categories in an instantaneous whole; the different modes are simultaneously active (for the speaker and the listener, who is trying to recreate the growth point). That one idea exists in two such different modes is the motive force for the utterance and its linked meaning formation.

4.2. Unpacking

The growth point is resolved by unpacking it into a more stable form, with a grammatical construction being the most stable outcome possible. Intuitions of good form (called 'intuitions-1,' the individual's direct perceptual experience of the static structure language) arise and are the *stop orders* for the dialectic. Once the speaker sensed a well-formed construction, she resolved the conflict by distributing the imagery and categorial content of the growth point into its prepared slots, and this stopped the dialectic process (how this might work is illustrated below). In this way, the dynamic intersects the static, as expected by Saussure's double essence insight—intersects it in fact in several ways: in the growth point, in the unpacking, and in the stop order. It is not that unpacking invariably reaches a full grammatical construction. It proceeds until some threshold of stability is reached, which may often be less than a complete outcome; or it may just break off if stability proves unattainable in the time spans attainable at socially realistic speech rates, e.g., because of an inappropriate construction attempt. Thus pauses and grammatical approximations, rife in daily discourse, can be explained as products of the dialectic resolution and the speaker's efforts toward it within limited time spans.

In this example, nonetheless, the growth point smoothly unpacked into a construction, the causative "someone drops (=causes to move by dropping) something down some landmark object." Intuitions-1 of the caused-motion construction arose and became the stop order, the construction plausibly resolving the dialectic by providing slots for the growth point image and its categorial content.

Subj	V	**Obj**	**Obl**
↕	↕	↕	↕
Ø (Tweety)	drops	**it (b-ball)**	**down**

(boldface for the slots that gathered the pieces of the growth point; the Tweety subject and the verb "drops" are explained below).

4.3. Context and fields of oppositions

Context is a second source of dynamism. Theoretically, a growth point is a psychological predicate in Vygotsky's [8] sense, a significant contrast within a specific context (cf. the concept of communicative dynamism [20]. While context reflects the physical, social and linguistic environment, it is also a mental construction; the speaker constructs this representation of context, in order to make the intended contrast meaningful within it. The growth point is thus not fixed and implies the context from which it is differentiated. Finding this context in actual data is an essential part of validating the growth point empirically. The mental construction of the context is modeled as a field of oppositions; what the speaker creates is a field of oppositions to make the psychological predicate differentiable within it. This is a model in which meaning is a *relationship* between a point of contrast and the background or field of oppositions from which it is being differentiated, not an accumulated 'substance'.

4.3.1. The catchment

A further concept, the catchment, provides an empirical route for finding this field of oppositions. A catchment comprises multiple gestures with recurring form features, and reveals the discourse segment to which the growth point belongs. More than one catchment can be simultaneously active for the same growth point. The full complement of catchments can suggest the oppositions from which the growth point is being differentiated.

To identify the catchment in the "it down' case, we look for other gestures in which the hands are shaped and/or move similarly to the target gesture, and see if these gestures comprise a family with thematic continuity. We find such a family; in the speaker's rendition, similar two-handed gestures had to do with the bowling ball in the role of an antagonistic force, contra-Sylvester.

4.4. The full description

We can thus further specify the "it down" growth point: it was a psychological predicate differentiating the bowling ball *as* this antagonistic force. Various antagonistic forces against Sylvester were the field of oppositions; the differentiated version was this force in the form of the bowling ball moving downward. The growth point and this context provide a richer picture of the

speaker's idea unit than a purely referential reading of the phrase, "drops it down the drainpipe," suggests:

> Ways of Thwarting Sylvester: Bowling Ball Downwards

Also, we can now explain the timing of the gesture: the downward thrust coincided exactly with the linguistic categorial content with which it formed a growth point, idea unit, or psychological predicate. It skipped the verb "drops," despite the fact that this verb described the bowling ball's motion down, precisely because the verb does not describe the bowling ball in its role as an antagonistic force; it describes what Tweety did, not the bowling ball, and thus could not have categorized the image with the intended meaning. The speaker's core idea was not dropping but the idea of the bowling ball moving down as an antagonistic force. Hence, the details of how gesture and speech combined, including timing, can be explained as aspects of the speaker's construction of the psychological predicate in the context, which is to say her thought process in context. (Other psychological predicates in the same catchment also conveyed the antagonistic force theme, specifying its effects on the unfortunate Sylvester: how he became a kind of living bowling ball, rolled down a street, into a bowling alley, and knocked over all the pins. Each of these can be analyzed in turn as psychological predicates differentiating further contrasts within the Antagonistic field of oppositions.)

The growth point was unpacked into a caused-motion construction, as noted, and we can analyze this and explain where the remaining pieces of the utterance, Ø (Tweety) and "drops", came from as well. Unpacking is more than just finding a construction in which to house a growth point; it includes the differentiation of further meanings each with their own contexts, and integrating them with the growth point so that the construction, including its semantic frame, can resolve it. The unpacking took place in a second catchment, also active during the speaker's representation of the bowling ball episode. The immediately preceding utterance was, "he tries going [**up**] the in**sid**][e of t**he drai**npipe]," which segued directly into our target utterance. The three gestures (in bold) were made the same way, with one hand rising upward, the first finger extended. Although this may include pointing, the gestures occurred with the theme of Sylvester acting as a force of his own (for many speakers an extended first-finger gesture conveys compression: Sylvester, inside the pipe, squeezes his plump body down to about half size). So, for this speaker, the utterance and the target utterance comprised a paradigm of opposed forces. Opposed forces was her way of construing the episode: not merely the bowling ball and Sylvester colliding, but Sylvester, a force moving up, versus the bowling ball, a force moving down—each force with its own gesture imagery. The bowling ball moreover was not the original antagonistic force; the sentence was "(Tweety) drops it down," which starts out with Tweety in the subject slot as the force. The speaker understood from the cartoon that she had to make the bowling ball into this force. The verb "drops" plus the caused-motion construction neatly achieved the shift from Tweety to the bowling ball. This is the growth point account how the verb and the Tweety subject made their way into the utterance.

The whole target utterance was thus the product of two contexts: 1) the growth point in the context of the bowling ball as an antagonistic force: this was the core idea unit; and 2) caused-motion with "drops" and Tweety as

subject: the further meanings in the paradigm of opposed forces that resolved the imagery-language dialectic, and shifted the antagonistic force to the bowling ball. The target utterance, although a single grammatical construction, grew out of two distinct contexts and gained oppositional meaning from each.

The linguistic side of a growth point is not necessarily grammatical. The "it down" growth point is not grammatical but nonetheless formed a growth point with the downward image in the context of thwarting Sylvester. Nor is it necessarily a verb (a popular psycholinguistic hypothesis that the verb is the starting point, is contradicted by the preparation phase passing straight through "drops"). The growth point can be any co-expressive linguistic category(ies) that enables the intended point of contrast to be differentiated within a field of oppositions built in part to make the contrast possible. Unpacking then must find a construction to resolve the growth point into a stable pattern.

Metaphoricity is present also. The downward moving bowling ball existed as something else, as an abstract idea of an antagonistic force. The importance of the metaphor is to enable the abstract, non-imaged meaning of an antagonistic force to become an image and to take part, as an image, in an imagery-language dialectic. In this way metaphoricity was an essential part of the growth point (not only in this case but in numerous others, perhaps all). This bowling ball metaphor was an impromptu creation but other gesture metaphors are culturally established but play the same role of enabling imagery-language dialectics with abstract unimageable meanings. An illustration is the 'palm up open hand', in which the hand(s) appear to present a discursive object. The metaphor is recognizable as the so-called 'conduit' metaphor, an image of the general metaphor culture (but not universal), in which an abstract idea is presented as if it were a substance in the hand or a container (cf. verbal examples like "the movie had a lot of meaning," where the movie is a container, or "she handed him that idea," where an idea is on the hand).[4]

4.5. Summary

The growth point is thus a theory of the cognitive core of utterances; what thought units are like as they begin; their incorporation of context; how they evolve dialectically and how imagery intersects linguistic form to create a surface utterance. However, it is a limited model. It says nothing of how growth points are activated. This includes lexical activation, as in the case study, where part of the categorial core – the "it" – was triggered by the ball reference (and word "ball") in the preceding clause. Models of lexical retrieval may apply but it is also possible that such models are inadequate to explain this kind of feedforward (since it is not actually 'feedforward' – the word "ball" was not just shipped ahead to become the next GP; rather it triggered a whole new precise idea in the speaker's mind, where the ball took on the role of antagonist). Also, we see in the "it down" case study that tracking the scope of recent co-references is assumed in the model but not explained – the "it"

[4]The gesture includes iconicity obviously, but also, in the placement of the hands in the upper central space, deixis indicating an upper space locus; and following Kevin Tuite's [21] idea that in every gesture there is a rhythmical pulse, something like a beat indicating that content has significance beyond its immediate setting, in the wider discourse, for example. Thus, one gesture includes all semantic components, and this is not a unique case. Multiple components is a reason for rejecting the idea of gesture types and thinking instead of dimensions—metaphoricity, iconicity, deixis, emphasis (i.e., beats), etc.—on which gestures load to differing degrees.

indicates co-reference vis-à-vis the earlier "ball" but there is no mechanism for this at present. It may be that some of the missing ingredients are matters of new elaborations (how the GP was initiated at the first mention of "ball" in the preceding clause for example, how "it" indexes the co-reference of the bowling ball), but others belong to another realm altogether – the proper modeling of speaker's purposes, for example, including the seemingly correct intuition that local purposes are created by the process of verbal thought as much as guiding them, beacon-like. Some of the lexical activation problems may be solved only once this further mystery is plumbed.

I apply this theoretical framework over a range of situations – discourse and gestures in different languages (Turkish, Spanish, Mandarin, as well as English); the gestures of children; the Whorfian hypothesis, arguing that the impact of language on imagery is often a dynamic effect concealed by the classic concentration in Whorf discussions on the static dimension but consistent with Slobin's thinking for speaking [22, 23]; linguistic impairments (aphasia; right-hemisphere damage, which impairs discourse cohesion; and the split-brain state, all of which were described in [3] but are now integrated into a new neurogestural model, in [1]).

5. Problems with modeling[5]

The global-synthetic property: this is the semiotic essence of gesture. Can it be modeled? Seemingly not, but read on.

The main sticking point for a computational model of the GP appears to be its character as a minimal dialectic unit. One aspect of the contradiction is the global character of imagery.

Global refers to the fact that the determination of meaning in a gesticulation proceeds in a downward direction. The meanings of the 'parts' of the gesture are determined by the meaning of the whole. In fact, parts come into being only in the meaning landscape of the whole; they have no independent existence (so, for example, the palms facing down mean agenthood, but the individual fingers meaning nothing; or in a different case, the first finger extended means compression but the palm means nothing – the parts depend in both cases on the global significances of their gestures). This semiotic model contrasts to the upward determination of meanings in sentences. *Synthetic* refers to the fact that a single gesticulation concentrates into one symbolic form distinct meanings that might be spread across the entire surface of the accompanying sentence.

The problem is that the use of features in computational models appears to force the process of gesture creation to be combinatoric, thus losing the opposition of semiotic modes essential to the dialectic (global imagery vs. combinatoric language). Features would be combinations of forms and meanings like: the hands a) facing down (→force downward), b) shaped around and over an imaginary sphere (→bowling ball), and c) moving downward jointly (→direction of bowling ball and force). In a model, such form-meaning pairs combine to create a gesture with the intended significance.

To be global, however, the process wants to work from the overall meaning downward. Even if we force a model to proceed in this direction, it

[5] I am grateful to the GP to the Max group at ZiF, the University of Bielefeld – Sue Duncan, Timo Sowa and Stefan Kopp – for freewheeling discussions of the material in this section.

appears that form features need to have their own meanings in order for a global meaning to find them – but do they? Here are some thoughts:

- The specific form features of the gesture are constrained by mechanical factors – where the hands already are, their current orientation, etc, which need not have anything to do with current significance.
- Suppose that significances trickle down into a configuration that already exists and Viv. (say) then improvises something that we, on analysis, decide means 'spherical', 'downward', and 'effort' – what does she need to do for this?
- She needs to perform *an action that embodies these meanings*. Does this imply combining form-meaning features? *Or is it enough to 'act'?* Is the action of propelling a bowling ball downward sufficient to generate a gesture with the significances we are after?
- The idea of *coordinative structures* (the Haskins-related action model) seems to apply, with the addition of a thought-language-hand link (accessing and steering coordinative structures using significances). Coordinative structures are not themselves significant forms; they are "flexible patterns of cooperation among a set of articulators to accomplish some functional goal" (anonymous Yale linguistics handout found by Google).

Using coordinative structures. The goal is to exploit the inherent flexibility of coordinative structures in such as way that significances activate and shape them. Do coordinative structures so managed avoid the combination problem or are they just a fig leaf? The question is: does the idea of the bowling ball as an antagonistic force moving downwards automatically take care of features such as size (largish), placement (upper), direction (down), and motive force (agenthood)? As I understand coordinative structures, they work like tuned springs. They start off from some initial state and tamp down as they approximate the target: an object or an image. If the attractor can be a real object, with a thought-language-hand link, as IW reveals, it can also be a *significance* (e.g., the idea of a bowling ball being thrust downward and its metaphoric meaning of an 'antagonistic force').

So, the resolution: ideas or significances are attractors of coordinative structures; the coordinative structures zero in on these attractors; the properties of the *attractor* bring out features in the coordinate structures interactively: so features are outcomes, not initial conditions, with significances that derive from the action as a whole, and this is the global property. There is no lexicon of feature-meaning pairs ('facing down →force downward' and the like). The features arise during the action itself. Once a gesture has been created it is usually true that we can identify features of form that carry meanings, but these are the outcomes of the gesture, not the source. Each coordinative structure is an 'action primitive', but the critical difference from a feature is that coordinative structures do not have significances. Cornelia Müller's implicit actions in gesture (drawing, outlining, sculpting, grasping, etc.) reemerge as packages of coordinative structure, or patterns of patterns adapted to objects, actions or shapes [15], now adapted to ideas, as kinds of metaphors at the origin of an imagery-language dialectic.

I can't judge the computational feasibility of this resolution, but it does seem to provide a way to generate global imagery with significances that descend from wholes to parts non-compositionally. It is conceivable, at least worth mentioning, that a hybrid analog-digital machine could correctly model

the growth point. The analog device itself could be simulated digitally, of course, but should simulate such properties as three-dimensional space, limited but varying granularity, differentiation of spatial blocks, orientations, etc. These properties establish the coordinative structures targeting significances, as described.

But not modeled: not modeled by coordinative structures is the growth point itself. Coordinative structures explain the global property, essential to a dialectic, but not the differentiation of psychological predicates; growth; inseparability from context; co-presence of imagery and linguistic categorization; the co-expressiveness of imagery and language; internal tension and motivation; or change/unpacking. In short, the 'essential duality' of language [9] of which the growth point is a minimal unit, seems at present impossible to model by a computational system.

6. Gestures and inhabitance

A further point places this entire discussion on a different plane and in so doing provides an answer to the question: what becomes of an imagery-language dialectic when gestures do *not* appear? We get a deeper understanding of the imagery-language dialectic by introducing the concept of a 'material carrier'. A material carrier is the embodiment of meaning in a concrete enactment or material experience. A material carrier appears to enhance the symbolization's representational power. The concept implies that the gesture, *the actual motion of the gesture itself*, is a dimension of meaning. Such is possible if the gesture *is* the very image; not an 'expression' or 'representation' of it, but is *it*. The gesture *itself* is a component of the dialectic. From this viewpoint, a gesture is an image in its most developed—that is, most materially, naturally embodied—form. The absence of a gesture is the converse, an image in its least material form. The greater the felt departure of being from the immediate context, the more likely its materialization in a gesture, because of its contribution to being. Thus, gestures are more or less elaborated depending on the importance of material realization to being. Absence of gesture is then the predictable result of a minimal departure from context; in repetitive or denatured contexts imagery fades and, Cheshire Cat-like, only the leer of imageless thought remains. Merleau-Ponty [24] expressed a similar view of language in *The Phenomenology of Perception*: "The link between the word and its living meaning is not an external accompaniment to intellectual processes, the meaning inhabits the word … What then does language express, if it does not express thoughts? It presents or rather it *is* the subject's taking up of a position in the world of his meanings" (p. 193). The "it down" growth point was this speaker's taking up of a position in the world of her cartoon narration, her momentary state of being, materialized in the image of the bowling ball as an antagonistic force.

7. The social/mental interface

A further dimension, as Vygotsky [8] famously argued, brings out that human thought is fundamentally social in character, even in the absence of an active interlocutor. Such implies that growth points are intrinsically social. The

growth point does not describe a mind-in-isolation. Social context effects were present even in the case study – that a gesture occurred at all presumed a listener, and the gesture was presented to the listener in central space. Any social minimalism reflects the limits of the circumstances, not a restriction of the concept itself, and in fact work in my lab in recent years, especially by my PhD students, has revealed the social-interactive context of the growth point.

Özyürek [25] showed that changing the number and the spatial loci of listeners has an effect on the speaker's gestural imagery. Thus, among the shaping factors in a field of oppositions was the speaker's social interactive context. Plugging this result into the growth point model, we infer that an imagery-language dialectic can be altered by changes of the social context.

And dialogues result in individuals inhabiting *similar growth points*. One can find two-party growth points, gestures from one person synchronizing with a second person's speech, and vice versa—someone's speech accompanied by another person's gestures (experiments by [26], [27], and Duncan pers. comm., respectively). Conversations are dynamically affected by the participant's gestures, even decisively altering direction when a conflict arises over the meanings metaphorized in the shared gesture space [28]. Such conflicts produce diverging imagery-language dialectics, which speakers attempt to realign.

On the other hand, when one speaker attempts to insert a false scene into a narration a joint GP is often impossible, as shown in the immediate breakdown of the interchange with the listener (the listener's confusion is the 'lie-detector'; research by Franklin). Finally, turn taking exchanges and interactions in group meetings can be explained in terms of 'mind-merging', in which turn-exchange signals synchronize GPs between outgoing and incoming speakers [29].

8. Language origins: 'the ultimate answer'

An important new source of observations is the case of IW, a man who suffered, as a young adult, sudden and complete deafferentation from the neck down [30]. IW relearned movement control by utilizing vision and cognition, and he controls motion in this way to perfection. He also performs gestures with speech synchrony and co-expressiveness and does so even without vision, a condition where nongesture instrumental actions are impossible for him. In other words, actions for IW organized by language and thought have properties beyond those of goal-directed actions. His case suggests a partial dissociation in the brain of the organization of gesture from the organization of instrumental action, and the existence of a dedicated thought-language-hand link that would be the common heritage of all humankind.

8.1. Evolution of the thought-language-hand link

We accordingly end with an attempt to provide 'the ultimate answer' to the question of an imagery-language dialectic—why it exists at all—with a theory of language evolution that focuses on this thought-language-hand link. I develop a hypothesis that the origin of language crucially depended at one point on gestures (I do not mean that the first form of language was gestural: I intend something quite different, as I will explain below). Without gestures, according to this hypothesis, the brain circuits required for language could not

have evolved in the way they apparently have. In common with much recent speculation, the theory presupposes the recently discovered 'mirror neurons', but adds something theoretical. This is 'Mead's loop' (named after the philosopher, George Herbert Mead [31], who wrote that "Gestures become significant symbols when they implicitly arouse in an individual making them the same response which they explicitly arouse in other individuals."), which I propose supplements the mirror neuron circuit.

8.1.1. Mirror neurons and Mead's loop

According to Mead's loop, what was selected in human evolution is a capacity, not present in other primate brains, for the mirror neuron circuit to respond to one's *own gestures* as if they belonged to someone else (thus gesture is activated as part of social interaction, producing among other things the social dependence of gestures when the addressee is invisible—speaking on the phone, a blind person talking to another blind person—but not speaking into a tape recorder). Crucially, Mead's loop brings the meanings of gestures into an area of the brain where actions are orchestrated. It provides a way for significances other than the significances of actions themselves to co-opt the action orchestration machinery of Broca's area, and explains how and under what conditions the IW-revealed thought-language-hand link could have evolved. A creature who possessed such a capacity, however minimally, would have had advantages in child rearing, for example, being better able to scaffold and error correct (plausible vectors for the origin of language being mothers and infant children; an origin of language *acquisition* as well as of language).

8.1.2. But not gesture-first

Contrary to a theory from Condillac enthusiastically resuscitated in recent years [e.g. 32] that the initial form of language was gesture, I am advocating that evolution selected an ability to *combine* speech and gesture; they had to occur jointly for the advantage to be realized. Speech and gesture would have evolved *together*. The plausibility of this hypothesis is enhanced by William Hopkins' observation that chimpanzees show hand dominance for gestures only when the movements co-occur with vocalization (Hopkins pers. comm.). The last common ancestor may therefore already have had a vocalization-gesture link. The thought-language-hand link could build on this precursor during its own selection via Mead's Loop. If there had also been a gesture-first step it would not have led to human language but to pantomime (pantomime could have its own evolution, landing at a different point on Kendon's Continuum, reflected today in different timing re speech—alternating rather than simultaneous). Just as speech could not have evolved without simultaneous gesture, gestures could not have evolved without a duet with speech [33].

References

[1] McNeill, David 2005. *Gesture and thought*. Chicago: University of Chicago Press.

[2] 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.

[3] McNeill, David. 1992. *Hand and mind: What gestures reveal about thought*. Chicago: University of Chicago Press.

[4] Mayberry, Rachel & Jaques, Joselynne. 2000. Gesture production during stuttered speech: insights into the nature of gesture-speech integration. In D. McNeill (ed.). *Language and gesture*, pp. 199-214. Cambridge: Cambridge University Press.

[5] Cassell, Justine, McNeill, David, & McCullough, Karl-Erik. 1999. Speech-gesture mismatches: evidence for one underlying representation of linguistic and nonlinguistic information. *Pragmatics & Cognition* 7: 1-34.

[6] Kelly, Spencer D., Barr, Dale J., Church, R. Breckinridge, & Lynch, Katheryn. 1999. Offering a hand to pragmatic understanding: the role of speech and gesture in comprehension and memory. *Journal of Memory and Language* 40: 577-592.

[7] Iverson, Jana M. & Goldin-Meadow, Susan. 1997. What's communication got to do with it? gesture in congenitally blind children. *Developmental Psychology* 33: 453-467.

[8] 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.

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

[10] Wundt, Wilhelm. 1970. The psychology of the sentence. In Arthur Blumenthal (ed. and trans.), *Language and P\psychology: Historical aspects of psycholinguistics*, pp. 20-33. New York: John Wiley & Sons Ltd.

[11] Hockett, Charles F. 1960. The origin of speech. *Scientific American*, 203: 88-96.

[12] Ishino, Mika. 2001. Conceptual metaphors and metonymies of metaphoric gestures of anger in discourse of native speakers of Japanese. In M. Andronis, C. Ball, H. Elston & S. Neuvel (eds.), *CLS 37: The main session*, pp. 259-273. Chicago: Chicago Linguistic Society.

[13] Reddy, Michael J. 1979. The conduit metaphor: a case of frame conflict in our language about language. In A. Ortony (ed.), *Metaphor and thought*, pp. 284-297. Cambridge: Cambridge University Press.

[14] Lakoff, George & Johnson, Mark. 1980. *Metaphors we live by*. Chicago: University of Chicago Press.

[15] Müller, Cornelia 2004. The palm-up-open-hand. A case of a gesture family? In C. Müller & R. Posner (eds.), *The semantics and pragmatics of everyday gestures*, pp. 233-256. Berlin: Weidler Verlag.

[16] Kendon, Adam 2004. *Gesture: Visible action as utterance*. Cambridge: Cambridge University Press.

[17] Bloom, Ralph 1979. Language creation in the manual modality: A preliminary investigation. Bachelors Thesis, Department of Behavioral Sciences, University of Chicago.

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

[19] McNeill, David 2000 Introduction. In D. McNeill (ed.), *Language and gesture*, pp. 1-10. Cambridge: Cambridge University Press.

[20] Firbas, Jan. 1971. On the concept of communicative dynamism in the theory of functional sentence perspective. *Philologica Pragensia* 8: 135-144.

[21] Tuite, Kevin. 1993. The production of gesture. *Semiotica* 93: 83-105.

[22] 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.

[23] 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.

[24] Merleau-Ponty, Maurice. 1962. *Phenomenology of perception* (C. Smith, trans.). London: Routledge.

[25] Özyürek, Asli. 2000. The influence of addressee location on spatial language and representational gestures of direction. In D. McNeill (ed.), *Language and gesture*, pp. 64-83. Cambridge: Cambridge University Press.

[26] Kimbara, Irene 2002. On gestural mimicry. Unpublished ms., University of Chicago, Department of Linguistics.

[27] Furuyama, Nobuhiro 2000. Gestural interaction between the instructor and the learner in origami instruction. In D. McNeill (ed.), *Language and gesture*, pp. 99-117. Cambridge: Cambridge University Press.

[28] McNeill, David 2003. Pointing and morality in Chicago. In S. Kita (Ed.), *Pointing: Where language, culture, and cognition meet*, pp. 293-306. Mahwah, NJ: Erlbaum.

[29] McNeill, David 2006. Gesture, gaze, and ground. In S. Renals & S.S. Benglo (Eds.), *MLMI 2005. LNCS 3869*, pp. 1-14.

[30] Cole, Jonathan. 1995. *Pride and a daily marathon*. Cambridge, MA: MIT Press.

[31] 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.

[32] Arbib, Michael 2005. From monkey-like action recognition to human language: An evolutionary framework for neurolinguistics. *Behavioral and Brain Sciences* 28: 105-124.

[33] McNeill, David, Bertenthal, Bennett, Cole, Jonathan, & Gallagher, Shaun 2005. Gesture-first, but no gestures? Commentary on Michael Arbib "From monkey-like action recognition to human language: An evolutionary framework for neurolinguistics. *Behavioral and Brain Sciences* 28: 138-139.