The intelligibility of gesture within a framework of co-operative action

Charles Goodwin
Applied Linguistics, UCLA

Gesture-first theories of language propose the transparent intelligibility of deictic and iconic gestures. The gestures of a man with a three-word vocabulary are used to investigate gesture without accompanying language. Rather than being transparent, the rich intrinsic meaningfulness of deictic and iconic gestures produces a surplus of possible referents. The task of working out their meaning delays movement to subsequent action, and thus creates selective pressure for the emergence of arbitrary, rather than inherently meaningful, signs. Analysis then turns to Kendon's argument that meaning and action are accomplished through the way in which talk, gesture, and phenomena in the environment mutually elaborate each other, with the semiotic possibilities of each of these resources mutually constraining the others.

Introduction

Kendon (2009) argues strongly that the phenomena human beings use to build action, including language and gesture, do not operate alone as isolated systems or modalities, but instead are organized within larger configurations of mutually interacting meaning-making resources. Here I will use the interactions of an aphasic man, Chil, who builds many of his utterances primarily through gesture, to probe the argument that language first originated in gesture, and then switched to speech. Chil's inherently meaningful, but ambiguous, gestures systematically delay the onward progression of the actions he is trying to accomplish through gesture. This suggests that the organization of action within interaction creates an environment that promotes the development of arbitrary signs. Consistent with Kendon's observations, rather than existing in a single modality, action is built by bringing together different kinds of semiotic phenomena, including arbitrary linguistic signs and gesture, into configurations where they can mutually operate

on each other. The specificity of an arbitrary sign constrains the possible referents a gesture might have. Such co-operations make possible rapid movement to a subsequent action.

Most discussion of such issues focuses largely on the actors' bodies (including not only gesture but also the rich frameworks for embodied co-orientation that Kendon has so eloquently described as grounding meaning through processes such as frame attunement (1990b; 1990a)) and language. However, the environment where action is situated is also massively attended to and implicated in the intrinsic organization of action. Interaction between scientists on an oceanographic ship will be used to extend the analysis of co-operative action to include both how subsequent action is built by performing structure-preserving transformations on the materials provided by a prior action and how these same practices build settings that accumulate through time resources which make possible the actions that occur within them.

Gesture-first theories of language origins

It has been argued by a number of different scholars that language evolved from a prior system in which meaning was displayed through gesture (Arbib 2005; Armstrong, Stokoe, & Wilcox 1995; Corballis 2002, 2012; Donald 1991; Tomasello 2008). Thus Corballis (2012: 206) argues that "there are no strong reasons to reject the notion that language was predominantly manual until comparatively recently in hominin evolution…" While engaging with such theories, and providing a forum for their presentation through his editorship of the journal *Gesture*, Kendon has eloquently articulated a skeptical stance towards them, arguing that a switch from gesture to spoken language never occurred since the primordial situation for language use is one in which both talk and consequential movements of the body work together (Kendon 2009).

Agreeing with Kendon, I will argue that gesture alone is not adequate for the production, or evolution, of the forms of meaning, and, more crucially, of co-operative action, that sit at the center of human language use. A major problem is that while both indexical and iconic gestures are inherently meaningful, they can, on specific occasions of use, refer to a plurality of candidate references (Goodwin 2011: 190–192). Tomasello proposes that the meaning of a gesture can be adequately constrained by the structures of mutual knowledge that constitute what Clark (1996) analyzes as common ground, within a distinctly human psychological infrastructure of shared intentionality (Tomasello 2008: 11). While recognizing the importance of such structures, I will use actions constructed primarily

through gesture by Chil, a man left with a three-word vocabulary after damage to the left side of his brain, to attempt to demonstrate that frameworks such as common ground do not constitute a general solution to the task of rendering gesture adequately intelligible for the needs of participants working to build action in concert with each other. Instead of moving forward immediately to the next action called for by the gestural utterance, Chil's interlocutors must first propose a gloss of what precisely was meant, which in turn must be accepted or rejected by a speaker building action almost entirely through gesture. Rapid movement to a relevant next action is systematically delayed by the need to establish what, out of many possibilities, this gesture refers to or proposes should follow next on this particular occasion of its use.

The way in which both deictic and iconic gestures are visibly saturated with inherent, but open-ended, meaning does not provide a clear path to language. Instead such seeable but potentially ambiguous meaningfulness constitutes a major obstacle that must be overcome. However, this very problem locates an environment lodged within the endogenous production of human action that creates selective pressure for the systematic evolution of arbitrary signs. By eliminating the necessity of working out, just after it occurs, what each new sign means and/or refers to, arbitrary signs make possible new forms of powerful, rapid co-operative action.

From a slightly different perspective, the human activity of building meaning and action in concert with others is not constituted within any single modality but instead brings together structurally different kinds of meaning-making resources that mutually elaborate and constrain each other (for example the visible orientation of an addressee towards the signs in whatever modality is being produced by a speaker or gesturer, within what Kendon calls a facing formation (1990a)). As argued by Streeck:

> Gestural understanding … is not the result of a shared grammar or lexicon, but of the coordinated embodied actions of people and their perspectives upon the material, real-world setting in which they interact. (2009: 4)

Gesture-first theories are inadequate because they confuse the modality of an action with the multimodal, multi-party interactive organization of co-operative action itself.

Chil and his resources

In 1979 Chil,[1] who had been a successful lawyer, a person who made his living through the use of language, suffered a stroke in the left hemisphere of his brain. For the rest of his life his spoken vocabulary consisted of three words: ***Yes, No*** and ***And***. Despite this he remained a powerful speaker in conversation. This was made possible in part by the fact that he retained highly expressive prosody, was able to gesture with one hand (including hand shapes for numbers), had excellent comprehension of what others were saying, and could respond gesturally and prosodically to what was being said with rapid fluency. By using these resources within the flow of ongoing interaction he was able to lead others to produce the words he needed. For analysis of Chil's abilities, and how he was able to make meaning in concert with others, see Goodwin (1995; 2003; 2004; 2006; 2007b; 2010; 2011).

The transparency of gesture?

> Human beings … find such gestures as pointing and pantomiming totally natural and transparent: just look where I am pointing and you will see what I mean. (Tomasello 2008: 1)

Because of Chil's inability to produce complex language he uses gesture extensively to build action directed towards others. Clearly he is not in the same situation as one of our ancestors acting within a world where language had not yet appeared. Chil not only understands language, but operates within sequences of action built through the language activities of his interlocutors. His use of gesture as a primary modality for displaying meaning does, however, provide a tragic natural experiment within which assumptions about the inherent transparency of gesture can be systematically examined.

It will be argued here that as a method for displaying meaning in the natural world, gesture is organized within a larger ecology of meaning-making practices. In light of this, Chil's gesture is not the same as that of a fluent speaker. Instead of co-occurring with rich speech, gesture alone must perform the primary work of showing someone what Chil wants to say. To accomplish this Chil, as seen in Figure 1, frequently joins multiple gestures together to build larger meaning-making packages. Indeed the gesture sequence here seems to have something

1. Chil was my father. Several years after his stroke I began to record his interactions at home when I visited him.

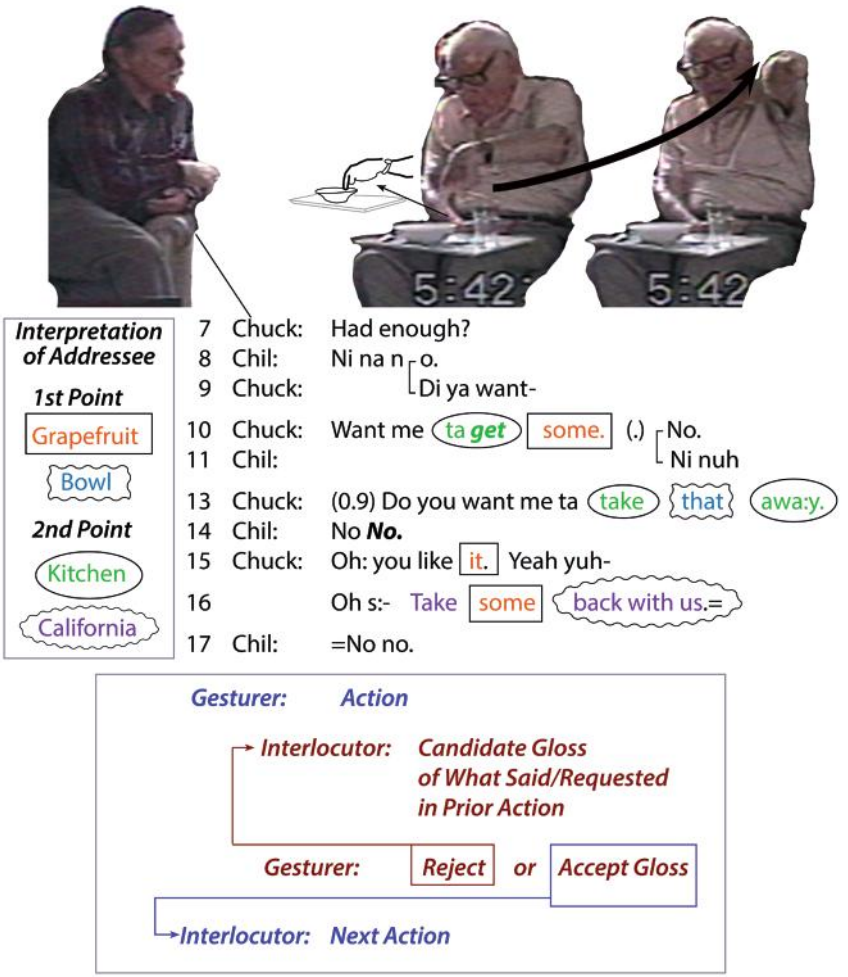

Figure 1. Addressee cannot recover either referent or action being done with linked pointing gestures

like a topic-comment structure. Chil first points to the bowl on his lap, where he has just eaten a special grapefruit sent from Florida. He then rapidly moves his pointing finger to some area in front of him, marking the second point as some way tied to the first. Such practices for combining and reshaping gesture so that it can function independently to display meaning is what would be required for a system in which action is built through gesture alone, e.g. the situation postulated by gesture-first theories. Among the spaces that fall within the trajectory of Chil's second point to the west are his kitchen, and California where his addressee, Chuck, lives.

Tomasello (2008: 224) notes within his discussion of how language might have emerged from gesture, other embodied practices, such as affective prosody and gaze towards another, which can contribute to the way in which action is built by gesture. This is central to Chil's interaction as well. Because of the length limitations of this chapter I have chosen as data a sequence where both Chil's prosody, and some aspects of his gesture use not central to the current argument, are investigated in more detail in other articles (Goodwin 2010; Goodwin 2011).

In Figure 1 lines 7, 8–10, and 13, Chil repetitively uses his two-part pointing gesture to build a series of actions.[2] Subsequent repetitions of the gesture package are responses to Chuck's inability to understand what Chil is trying to tell him. Despite repeated work Chuck is unable to figure out either what the gestures are pointing at, or what next action they are asking Chuck to perform.

As demonstrated by his addressee's inability to locate either what Chil is pointing at, or what action he is attempting to perform with the gestures, Chil's pointing gestures are in no way transparent. Both pointing gestures (and iconic ones) can indicate a range of different referents. An addressee not being provided with further specification, for example through co-occurring language, is faced with the task of uncovering which possible referent is the appropriate one for the action currently in progress. Chil's first point provides a particularly strong demonstration of this. His finger is almost touching the bowl beneath it. However, Chuck is unable to determine whether he is pointing towards the grapefruit (lines 10, 15, & 16), or towards the bowl (line 13). Chil's guess about the referent of the second point moves over 2,000 miles from Chil's kitchen in New Jersey to California between lines 13 and 16.

Several minutes later Chuck learns that his wife Candy had been walking a dog on the street in back of the house, another space that fell within the scope of Chil's second point. They then work out together that Chil wanted Chuck to offer some of the grapefruit to Candy.

It might be argued that gesture-first theories can accommodate such lack of transparency. In Tomasello's (2008) analysis the pointing gestures of humans differ in most significant ways from those of apes. First, they are embedded within a complex psychological framework founded upon shared intentionality. Central to this is the ability of human beings to recognize communicative intentions. That is certainly the case in in Figure 1. Chuck produces his candidate understandings of what Chil's gesture might mean (lines 7, 10, 13, 15, 16) precisely because he recognizes that Chil is trying to tell him something with the gesture, and indeed lead him to ***do*** something. However, recognizing the presence of a communicative intention within the action is not nearly enough. To go further Tomasello draws

2. See Goodwin (2011: 187) for a more detailed transcript.

upon Clark's (1996) analysis of common ground, knowledge that is shared by both speaker and addressee, and which provides for the intelligibility of actions that assume such knowledge. Tomasello (2008: 2) uses a series of anecdotes in which a gesturer walking with a female colleague points towards a bike outside a university library to demonstrate how common ground provides interpretative frameworks to make visible different kinds of action being done through the pointing. If the addressee has just broken up with her boyfriend and they both know this is his bike, the point might be warning her that he is in the library. However, if the bike has been stolen the point might indicate that it has been found, etc.

It would be difficult to find interlocutors who share more common ground than Chuck and Chil. Chuck is Chil's son and they have known each other intimately for over 50 years. Though they now live on opposite sides of the American continent, Chuck has just spent several days with Chil. Most importantly, he has been sitting with Chil during the entire time he was eating the grapefruit: the crucial interpretive frame for the pointing that occurs in Figure 1. The one crucial thing that Chuck doesn't know is that Candy is now walking behind the house. A system that would provide for the understanding of pointing by requiring that the contents of actors' heads be supplied with everything they need to know to understand the point before it occurs (e.g., that Chuck know where Candy is at this moment) does not work. This is not to deny the general relevance of common ground, but to demonstrate that something else is required to systematically account for the intelligibility of mundane pointing.

Action consequences of the indeterminacy of gesture

Chuck is not simply trying to recognize what Chil is pointing at. Instead he is trying to work out what ***action*** Chil wants him to perform next. His changing proposals about the referents of Chil's points are embedded within alternative possibilities for future action (e.g. taking the bowl away – to be cleaned in the kitchen – line 13 vs. bringing grapefruit back to California in line 16).

Such phenomena demonstrate the centrality of an action framework, rather than merely a referential one, for posing the issue of how gesture is intelligible. Chuck is being asked to do something now. He focuses on Chil's gestures as resources that might enable him to figure out what that action is.

More generally, I have argued (Goodwin 2012) that human beings build action co-operatively by performing structure-preserving transformations on materials placed within a public environment by others. Participants' actions are co-operative in that each is building next actions by performing systematic operations on materials provided by the other. Thus, to determine what he should do next Chuck

operates on each of Chil's points by transforming them into possible locations, displaying them through language rather than gesture. By decomposing Chil's action-complex into separate parts, and making use of each point by transforming it into something relevant to a possible next action, Chuck's actions preserve with modification the structural materials made available to him by Chil. When Chil uses variants of "No No" to reject each of Chuck's proposals, he in turn indexically incorporates what Chuck has just said into the intrinsic organization of his own action. Thus in line 14 Chil is not heard to be using "No No" as an isolated self-contained statement, but instead to be saying that he doesn't want Chuck to "take that awa:y." Action emerges through accumulative co-operative transformations on a progressively changing public substrate that at each iteration becomes the new point of departure for subsequent action.

Co-operation is central to the organization of human action not only on larger scales such as food sharing, but as a constitutive feature of the practices used to build local action in concert with others within individual utterances. Moreover, through the way in which developing action reuses structure that is also being used by interlocutor(s), this process progressively accumulates a body of shared resources and knowledge as public practice in ways that may be relevant to the constitution of a common ground. In addition to a psychological infrastructure that makes possible communicative intentions, analysis must also focus on the actual co-operative practices through which the materials used to build action, including gesture, are understood in just the ways that make possible subsequent action.

Gestures enter the arena for action endowed with rich but indeterminate visible meaning (*something* is being brought to the addressee's attention through the act of pointing). Their surplus of possible meaning has strong structural consequences for the ongoing organization of co-operative action. Chuck is faced with the task of operating on Chil's gestures to build his own response to whatever they might be asking him to do. However, he does not yet know what precisely these gestures are indicating or proposing. Instead of moving immediately to the (unknown) action requested by Chil, Chuck's next action takes the form of a candidate gloss, a tentative rephrasing in words of what Chil might be asking him to do (see the diagram under the transcript in Figure 1). Another action then ensues in which Chil either accepts or rejects this gloss. If the gloss is rejected, as it is repetitively here, the parties are unable to move forward to whatever action Chil is proposing should happen next. The production of glosses as understanding checks is pervasive in interaction with Chil (Goodwin 1995; 2003; 2004).The forward movement of action built primarily through gesture is stalled by the work of publicly establishing what these gestures mean as points of departure for subsequent action.

Co-operative action as an environment promoting the evolution of arbitrary signs

> Words are for the most part arbitrary symbols, bearing no natural relation to objects or events in the real world. The idea that language originated in manual gestures helps overcome this problem, since gestures can be shaped to represent what they refer to. (Corballis 2012: 203)

For Corballis the arbitrary nature of most linguistic signs poses a problem that is solved by the inherent meaningfulness of gesture. What was seen in Figure 1 suggests just the opposite. Within the framework of co-operative action, arguably the most pervasive environment for the emergence of both gesture and language in the natural world, the task of building a relevant next action is posed continuously as each current action comes to completion. Building action with inherently meaningful, but referentially ambiguous, signs, such as Chil's pointing gestures, retards fluent forward movement since participants must stop to publicly establish how these gestures are to be interpreted on this particular occasion of their use.

Signs that are arbitrary, rather than inherently meaningful, provide a powerful solution to this problem. If Chil could have said "Let's give Candy some of this delicious grapefruit" or even "Candy," none of the extensive work these participants perform together to try to publicly make visible what Chil is asking Chuck to do would have been necessary. Precisely because of their lack of resemblance to what they represent the relevant meaning of arbitrary signs is established immediately by convention, instead of opening up an arena for protracted interpretation that must be publicly validated through further collaborative interaction. Such issues emerge acutely within a framework where participants continuously face the task of rapidly building a next action that uses in detail the structure provided by a prior action as its point of departure.

Consistent with Kendon (2009: 363), I would argue that what would seem to be at issue here is not a switch from one embodied modality, such as gesture, to another, such as spoken language. Instead a crucial transition involves the emergence and pervasive use of arbitrary signs. Sign languages demonstrate that this can occur as effectively with manual signs as with spoken ones. Indeed, I expect that much of the most enlightening research on how the distinctive signs that constitute language arise will come from the study of emerging sign languages (Kegl, Senghas, & Coppola 1999; Haviland, this volume; Sandler, Meir, Padden, & Aronoff 2005).

What has been seen here suggests that the task of producing co-operative action in concert with others might provide an environment that systematically promotes the gradual evolution of practices for building meaning that use

arbitrary signs, instead of being restricted to inherently meaningful ones. One would not replace the other. As current human interaction richly demonstrates, action flourishes within a semiotic ecology in which arbitrary signs and gesture mutually elaborate each other to produce meaning and action that would be impossible for either in isolation.

Building action within an ecology of meaning-making practices

> Speech was not invented to overcome the disadvantages of signing. … A much better approach, it seems to me, and one that takes into consideration how utterances are actually produced in modern speakers, would be to start with the assumption that the transition into referential or language-like expressions involved hands and body, face and voice and mouth, all together, as an integrated ensemble. What so many writers on this topic – "gesture firsters" and "speech firsters" both – pay little attention to is the fact that modern humans, when they speak together in face-to-face situations, especially in the informal settings of everyday interaction, always mobilise face and hands and voice together in complex orchestrations. (Kendon 2009: 363)

The difficulties Chil has with the interpretation of his pointing arise from his inability to accompany his gestures with rich language, that is, as Kendon states above, to build action in an environment where "hands and voice work together in complex orchestrations." Working in isolation his gestures are characterized simultaneously by a surplus of meaning and an essential incompleteness.

I now want to briefly examine how action is built by bringing together meaning-making practices with very different properties, including pointing gestures and arbitrary language signs, where each can mutually elaborate the others.

Figure 2 depicts a sequence of action that occurred on an oceanographic research vessel studying the processes that occur as the Amazon River flows into the Atlantic Ocean. Investigating the talk and pointing that occurs here requires some understanding of what the participants are doing. A probe called a **CTD** (for **C**onductivity, **T**emperature and **D**epth) has been lowered into the ocean. The CTD contains sensors and bottles that can collect water samples at different depths. **Phy**llis, a **Phy**sical Oceanographer, is responsible for guiding the CTD through the water column. What we can call a sonar chart produces a complex image that allows someone with the proper professional vision to see how the CTD is positioned with respect to the bottom (the image on the bottom right of Figure 2 depicts the ship with the CTD under it, not the marks on the sonar chart). The CTD is expensive and there is a real danger of losing it if it gets caught in the mud that gets progressively thicker as the sea floor is approached. However,

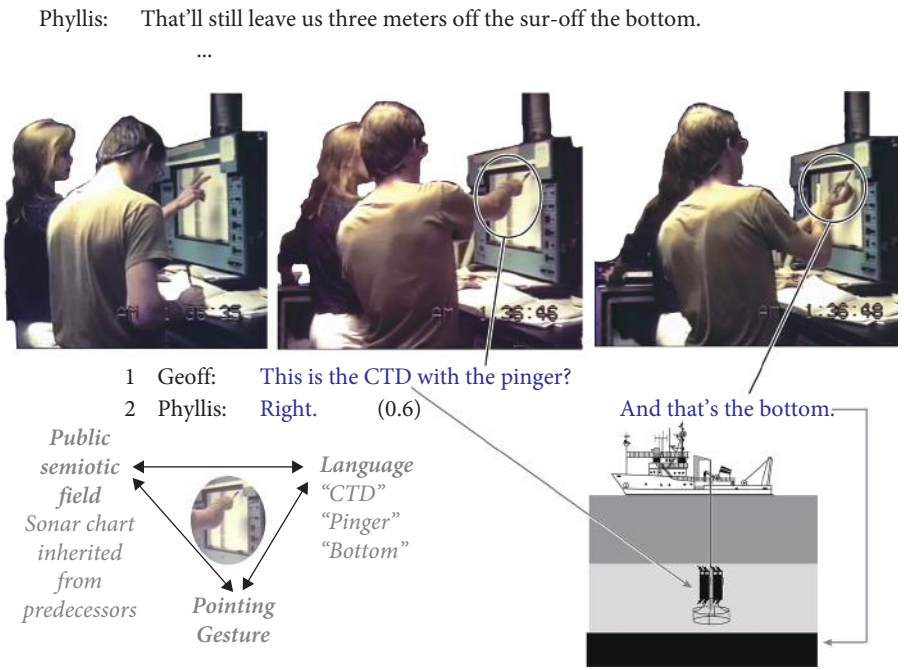

Figure 2. Co-operative action encompassing talk, pointing, structure in the environment, and the setting

for scientific purposes it is desirable to get as close to the bottom as possible. As Phyllis tries to determine how far she can lower the CTD, she repetitively points towards a particular area on the top right of the sonar chart, touches it, and occasionally makes notes at this place on the chart. She sometimes talks out loud while looking at the chart about the issues she is dealing with, while staring intently at the chart: "That'll still leave us three meters off the sur- off the bottom."

Geoff on the right is a **Geo**chemist using the CTD to collect water samples. He does not know how to interpret the complex images on the sonar chart. However, all of Phyllis' talk and pointing at the chart is accessible to him, though he is not in any way being addressed. What he has seen and heard enables him to make inferences about the patterning visible on the chart. In line 1 he uses his pencil to point at a particular place on the chart in the region where Phyllis has been pointing and asks if "This is the CTD with the Pinger?"

Both his pointing and what he says are only intelligible if participants, as well as analysts, take into account a complex ecology in which a range of very different kinds of meaning-making practices (language structure, pointing gestures, and the patterning visible on the sonar chart) are brought together to build action

by mutually elaborating each other, i.e. co-operatively (Goodwin 2012). If Geoff were to point without saying anything he would create a situation quite similar to that found when Chil points. Interlocutors would be faced with the task of making guesses about what he might be indicating and trying to do so through such a point. Instead, the arbitrary signs in his co-occurring talk constrain the range of interpretative possibilities offered by the pointing gesture alone. On the other hand the talk in isolation, with its deictic organization, cannot be adequately understood without the interlocutors also attending to the gesture and the phenomena being pointed at, the sonar chart. What we find here is quite consistent with Kendon's argument that "hands and body, face and voice and mouth, all [work] together, as an integrated ensemble" (Kendon 2009: 363), and that this is a relevant environment for asking how talk and gesture might have evolved together as humans progressively developed practices for building action and constituting a shared intelligible world in concert with each other.

However, in addition to language and gesture something else must be added to this complex of mutually elaborating meaning-making domains: the structured environment, here the sonar chart that is the focus of the participants' work, attention, and pointing. To ignore the environment that the participants are attending to, and focus only on their talk, gesture, and other embodied actions, is to draw an invisible analytic boundary at the skin of the actors. With such a boundary in place, crucial aspects of what they are doing together become inaccessible to analysis. The world being pointed at is frequently described in discussions of pointing (for example Tomasello's (2008) descriptions of bikes outside libraries, my own earlier discussion of the different kinds of phenomena that might be indicated by Chil's pointing, etc.) but rarely thematized as a topic for analytic focus. Yet participants themselves pay great attention to the task of rendering the world that is the focus of their action intelligible to each other in precisely the ways that will make possible the activities they are carrying out together. The images on the sonar chart are messy and complex; I was never able to read them. However, in that they provide structure-preserving transformations of the world under the sea that is being acted within, understanding these marks in enough detail to avoid losing the CTD in the mud is a crucial skill for someone who is manipulating it.

How might such professional vision be acquired through public practice? Elsewhere I have argued that environmentally coupled gestures (Goodwin 2007a), which link pointing, language structure, and the phenomena in the environment being pointed at (see the diagram at the bottom left of Figure 2), are powerful resources for constructing competent members of endogenous communities, such as professional archaeologists or surgeons. The mutually elaborating organization of environmentally coupled gestures brings together crucial categories, such as

"the bottom" or "the CTD," with actual examples of the complex phenomena in the environment, such as the squiggles on the sonar chart, that are to count as proper instantiations of those categories. Through environmentally coupled gestures abstract knowledge is linked to the embodied skill and professional vision required to act as a competent member of a community, that is, to know and understand the environment that is the focus of their attention in the ways that make possible relevant work within it.

In line 2 Phyllis builds a response to Geoff by performing structure-preserving transformations on not only his talk, but also his gesture, and the environmental field that was indicated through his gesture. Her talk begins with "Right" which indexically incorporates what he has just said and shown her. She then uses the structure of his utterance as a template for the organization of her own. His "This is" is preserved with transformation as "that's" and her "the bottom" occupies as a categorical alternative the position filled in his utterance by "the CTD with the Pinger." She matches his pointing gesture with one of her own (note the overlapping hands in the third image). Just as he located a specific place on the chart so does she. Her response is built as an ensemble of structure-preserving transformations that demonstrated her precise orientation to (1) the details of his talk; (2) his embodied action; and (3) structure in the environment that was the explicit focus of his gaze, pointing gesture, and probing query. It is not any single field in isolation, such as language or gesture, that is focused on by the participants in the organization of their action, but instead this entire complex of mutually elaborating semiotic fields.

Co-operative action and accumulative settings

The intrinsically co-operative organization of the actions found here is manifested in a number of different ways. First, individual actions are created through the intersection of different kinds of semiotic fields, such as gesture, structure in the environment, and arbitrary language. The way in which each operates on the others creates an emergent whole not found in any field in isolation. Building action by combining structurally different kinds of materials makes possible intrinsically multi-party action: a hearer can contribute simultaneously to the organization of an utterance in progress through gaze and other embodied displays, and stance can be marked prosodically and by facial displays of hearer as well as speaker (Goodwin 1980). Indeed, despite his three-word vocabulary Chil can make complicated statements laminating his prosody on complex language produced by others (Goodwin 2012). Because of the way in which actions are constructed through

the mutual elaboration of different kinds of semiotic fields, they are organized as a dynamic ecology, one capable of creative re-arrangement to adapt to significant disturbances. Thus typically gesture is explicated with co-occurring talk by the party producing the gesture. Chil's three-word vocabulary makes it impossible for him to produce relevant language structure as he gestures. However, as was seen in Figure 1, his interlocutors can provide the required language structure through candidate glosses. The mutually elaborating relationship between talk and gesture is preserved through adaptive changes in the participant producing the relevant talk, and the sequence required to accomplish that.

Second, as demonstrated by Phyllis' action-complex in line 2, subsequent action can be built by re-using with transformation the resources provided by earlier action. Rather than standing alone in splendid isolation, Phyllis' utterance is built co-operatively with Geoff's, as she incorporates with modification meaning-making resources he has provided into the organization of her own action (e.g. her parallel point to the chart, though landing at a different place and accompanied with a different description).

Such systematic accumulation of relevant structure provided by predecessors provides cognitive and practical organization for the setting itself (Hutchins 1995). Both the tools of the oceanographers, and the ship that carries them, incorporate work-relevant solutions to recurrent, systematic problems found by earlier actors: winches that can move heavy equipment such as the CTD from the deck of a ship to sea, sonar and the computer displays produced by the CTD that can provide an image of relevant structure in the sea beneath the ship, Niskin bottles that can collect seawater at different depths, etc.

This setting constitutes an accumulation of knowledge and resources that make possible the accomplishment of consequential action. These are clearly relevant to the role played by common ground in the analysis of Clark (1996) and Tomasello (2008). However, though it includes shared knowledge, a setting exists as a public landscape, a historically sedimented world. Newcomers must learn to navigate through settings by using the specific resources that are found within them to build consequential action, and thus become appropriate members: competent practitioners of the activities the setting makes possible. The kitchens found in our homes provide a mundane example (Goodwin 2009). One knows the world by acting within it, and this process includes not only the work being done by our co-participants, but also the contributions of our predecessors.

Conclusion

A number of different theorists have proposed that language might have evolved from an earlier stage in which meaning and action were accomplished largely through gesture (Arbib 2005; Armstrong et al. 1995; Corballis 2002; Corballis 2012; Tomasello 2008). Kendon (2009) has argued instead that for modern speakers talk and gesture work together, and that there is no reason to suppose that this situation was different for our ancestors. The phenomena of analytic interest are not organized within discrete modalities, and the proposed switch from one modality to another fails to take into account their interdependent organization. Within gesture-first proposals a main advantage of gesture is argued to be the way in which it clearly and naturally conveys meaning.

To probe gesture-first theories I examined interaction in which a man, Chil, left with a three-word vocabulary after a stroke, used gesture packages as a primary resource for constructing utterances to perform relevant action. Despite very close proximity between his pointing finger and what was being pointed at, the meaning of his pointing gestures was in no way transparent. Though clearly having reference, a particular gesture could in fact refer to a range of different phenomena on actual occasions of use (the same is true for his iconic gestures, but investigating these was beyond the scope of this chapter). Moreover, his addressee was his son whom he had known for over 50 years, and who had just been interacting with him. The parties shared very strong common ground.

The distinctive shape taken by interactions with Chil suggested that the ability to produce action in concert with others, rather than representation through gesture in isolation, should be a primary analytic focus. The inability of Chil's gestures to unambiguously convey meaning systematically led to delays in movement towards the next action he was requesting with his gestures. Faced with the inherent ambiguity of the signs he was producing, interlocutors had to present guesses about what he was trying to say, which he then had to accept or reject, so that the fluent production of action sequences was systematically retarded. Rather than providing a clear route to language, the inherent meaningfulness of gesture created an obstacle to be overcome.

Arbitrary signs offer a solution to this problem. It was proposed that the organization of co-operative action, the ability of participants to construct changing action in concert with each other that was rapid, efficient, and flexible by bringing together different kinds of meaning-making resources, creates an environment that would promote the evolution of arbitrary signs. Co-operative action packages that included arbitrary signs could incorporate the genuine representational power of gesture, while constraining its ambiguity through mutual elaboration with unambiguous signs.

Interaction on an oceanographic research ship was then used to investigate in more detail how pointing is organized within a larger semiotic environment that includes not only co-occurring talk, as emphasized by Kendon, but also consequential structure in the environment that is the focus of the participants' activities and which is invoked in a relevant fashion through their pointing gestures. The organization of co-operative action is central to this process on many different time scales. Simultaneously, at single moments in time, action is built by assembling resources, such as arbitrary signs in language, pointing gestures, and consequential structure in the environment, such as the marks on a sonar chart being pointed at. Sequentially, a next action was built by performing structure-preserving transformations on many or all of the varied components of the action being responded to, including its arbitrary signs, the use of pointing, and the structure of the environment being pointed at. On a much larger time scale the accumulation with transformation of materials put in place by earlier actors building action in the past creates dense settings and resources, such as the sonar chart, the varied equipment in their lab, the winch used to move their instruments to the sea, etc., that provide the resources used to build current action. Participants inhabit and act within a landscape of knowledge, tools and resources created through the accumulative work of their predecessors. Because of this a pointing gesture accompanied by relevant language works to constitute and index very rich, culturally specific professional vision: demonstrating to someone how to interpret complex squiggles on a chart, that simultaneously mark the positioning of distant objects being used to accomplish the work that defines the activities of their profession. Such arrangements of distinctively human action built through the mutual elaboration of many different kinds of resources (Kendon 2009: 363) provides the matrix for language, gesture, and the ability of human beings to operate together on complex worlds that they progressively change through these same actions.

Acknowledgments

I am deeply grateful to Michael Smith, David DeLiema, Candy Goodwin, and Pauline Beaupoil for most insightful comments on an earlier version of this chapter, and most especially to Chil and his family, and to Willard Moore for making fieldwork on the oceanographic vessel possible.

References

Arbib, M. A. 2005. "From monkey-like action rocognition to human language: An evolutionary framework for neurolingujistics." *Behavioral and Brain Sciences* 28: 105–168.

Armstrong, D. F., Stokoe, W. C., and Wilcox, S. E. 1995. *Gesture and the Nature of Language*. Cambridge: Cambridge University Press. DOI: 10.1017/CBO9780511620911

Clark, H. 1996. *Using Language*. Cambridge: Cambridge University Press. DOI: 10.1017/CBO9780511620539

Corballis, M. 2002. *From Hand to Mouth: The Origins of Language*. Princeton, NJ: Princeton University Press.

Corballis, M. C. 2012. "How language evolved from manual gestures." *Gesture* 12 (2): 200–226. DOI: 10.1075/gest. 12.2.04cor

Donald, M. 1991. *Origins of the Modern Mind: Three Stages in the Evolution of Culture and Cognition*. Cambridge, MA: Harvard University Press.

Goodwin, C. 1995. "Co-constructing meaning in conversations with an aphasic man." *Research on Language and Social Interaction* 28 (3): 233–260. DOI: 10.1207/s15327973rlsi2803_4

Goodwin, C. 2003. "Conversational frameworks for the accomplishment of meaning in aphasia." In *Conversation and Brain Damage*, C. Goodwin (ed.), 90–116. Oxford, New York: Oxford University Press.

Goodwin, C. 2004. "A competent speaker who can't speak: The social life of aphasia." *Journal of Linguistic Anthropology* 14 (2): 151–170. DOI: 10.1525/jlin.2004.14.2.151

Goodwin, C. 2006. "Human sociality as mutual orientation in a rich interactive environment: Multimodal utterances and pointing in aphasia." In *Roots of Human Sociality*, N. Enfield and S. C. Levinson (eds), 96–125. London: Berg Press.

Goodwin, C. 2007a. "Environmentally coupled gestures." In *Gesture and the Dynamic Dimension of Language*, S. Duncan, J. Cassell, and E. Levy (eds), 195–212. Amsterdam: John Benjamins.

Goodwin, C. 2007b. "Interactive footing." In *Reporting Talk: Reported Speech in Interaction*, E. Holt and R. Clift (eds), 16–46. Cambridge: Cambridge University Press.

Goodwin, C. 2009. "Embodied hearers and speakers constructing talk and action in interaction." *Cognitive Studies* 16 (1): 51–64.

Goodwin, C. 2010. "Constructing meaning through prosody in aphasia." In *Prosody in Interaction*, D. Barth-Weingarten, E. Reber, and M. Selting (eds), 373–394. Amsterdam: John Benjamins. DOI: 10.1075/sidag.23.29goo

Goodwin, C. 2011. "Contextures of action." In *Embodied Interaction: Language and Body in the Material World*, J. Streeck, C. Goodwin and C. D. Lebaron (eds), 182–193. Cambridge: Cambridge University Press.

Goodwin, C. 2012. "The co-operative, transformative organization of human action and knowledge." *Journal of Pragmatics* 46: 8–23. DOI: 10.1016/j.pragma.2012.09.003

Goodwin, M. H. 1980. "Processes of mutual monitoring implicated in the production of description sequences." *Sociological Inquiry* 50: 303–317. DOI: 10.1111/j.1475-682X.1980.tb00024.x

Hutchins, E. 1995. *Cognition in the Wild*. Cambridge, MA: MIT Press.

Kegl, J., Senghas, A., and Coppola, M. 1999. "Creation through contact: Sign language emergence and sign language change in Nicaragua. In *Language Creation and Language Change: Creolization, Diachrony and Development*, M. DeGraff (ed.), Cambridge, MA: MIT Press.

Kendon, A. 1990a. "Behavioral foundations for the process of frame-attunement in face-to-face interaction." In *Conducting Interaction: Patterns of Behavior in Focused Encounters*, A. Kendon (ed.), 239–262. Cambridge: Cambridge University Press.

Kendon, A. 1990b. "Spatial organization in social encounters: The F-formation system." In *Conducting Interaction: Patterns of Behavior in Focused Encounters*, A. Kendon (ed.), 209–238. Cambridge: Cambridge University Press.

Kendon, A. 2009. "Language's matrix." *Gesture* 9: 355–372. DOI: 10.1075/gest. 9.3.05ken

Sandler, W., Meir, I., Padden, C., & Aronoff, M. 2005. "The emergence of grammar: Systematic structure in a new language. *Proceedings of the National Academy of Sciences* 102 (7): 2661–2665. DOI: 10.1073/pnas.0405448102

Streeck, J. 2009. *Gesturecraft: The Manu-Facture of Meaning*. Amsterdam: John Benjamins. DOI: 10.1075/gs.2

Tomasello, M. 2008. *The Origins of Human Communication*. Cambridge, MA: MIT Press. A Bradford Book.