On the Human "Interaction Engine"

Stephen C. Levinson

My goal in this chapter is to make the case that the roots of human sociality lie in a special capacity for social interaction,[1] which itself holds the key to human evolution, the evolution of language, the nature of much of our daily concerns, the building blocks of social systems, and even the limitations of our political systems.

Much of the speculation about the origins and success of our species centers on the source of our big brains, the structure of our cognition, on the origins of language, the innate structures that support it, and on the striking cooperative potential in the species. These are genuine and important puzzles, but in the rush to understand them, we seem to have overlooked a core human ability and propensity, the study of which would throw a great deal of light on these other issues. It is right under our noses, much more accessible than the recesses of our brains or the fossils that track our evolutionary origins, and quite understudied. It is the structure of everyday human interaction.

Despite the fact that it is over fifty years since human interaction was first treated as a scientific object of inquiry deserving of a natural history (Bateson 1955; Chapple and Arensberg 1940; see also Kendon 1990), progress has been quite limited. One problem has simply been that human interaction lies in an interdisciplinary no-man's land: it belongs equally to anthropology, sociology, biology, psychology, and ethology but is owned by none of them. Observations, generalizations and theory have therefore been pulled in different directions, and nothing close to a synthesis has emerged. In this chapter, I therefore try to stand back and extract some generalizations about the special human abilities that seem to lie behind the structure of social interaction.

Are there Special Principles of Human Interaction?

Human interaction, by comparison with what goes on in even our nearest relatives, looks very distinctive, suggesting that there may be specific principles or abstract properties that underlie it. One starting point would be to ask whether there is a core universal set of proclivities and abilities that humans bring, by virtue of human nature, to the business of interaction—properties of interaction that are at source independent of variations in language and culture. Although much might be attributable to language, there are quite good prima facie grounds for thinking that human interactional abilities are at least partially independent of both language and culture:

■ Travelers to foreign lands report successful transactions conducted without language. Captain Cook's unintended sojourn in Cape York is a case in point, or Thomas Henry Huxley's journeys on HMS *Rattlesnake*. The best documentary evidence is probably the film *First Contact* (Connolly and Anderson 1987), incorporating footage made by the gold prospecting Leahy brothers contacting tribes in Highland New Guinea for the first time in the 1930s: it is as if the basis for transactional interactions exist independently of culture and language, and the slots can in necessity be filled by mime and iconic gesture (see Goodwin this volume).[2] Quine's (1960) demonstration of the impediments to "radical translation" notwithstanding, something like it seems anyway to occur.

■ Infants show an early appreciation of the give and take of interaction (Bruner 1976) long before they speak, indeed arguably at four months (Rochat et al. 1999), only two months old (Trevarthen 1979), or even 48 hours (Melzoff and Moore 1977), depending on the measure. By nine months old, infants are embarked on complex triadic interactions between ego, alter, and an object in attention (Striano and Tomasello 2001). We know that different cultures have different infant-caretaker patterns (see Gaskins this volume), so it is hard to rule out early cultural influence, but the infant evidence is highly suggestive of an ethological basis on which cultures may or may not choose to build in early infancy.

■ When language is lost, interaction doesn't disappear—restricted channels of communication, as in aphasia, can nevertheless support rich interaction (Goodwin 2003).

■ There is some evidence for a distinct "social intelligence" (Gardner 1985; Goody 1995) from inherited deficits and neurological case

studies. The study of autism and Asperger's syndrome, in comparison with, say, Down's syndrome kids, suggest a double dissociation: high-reasoning abilities, low social skills (Asperger's), low-reasoning skills, high social skills (high-functioning Down's)—see Baron-Cohen 2000 and Baron-Cohen et al. 1985. Similarly, different kinds of frontal lobe lesions induce different kinds of interactional incompetence, for example right temporal lesions correlate with flat affect and the loss of nonsuperficial understandings as required for jokes (Kolb and Whishaw 1990:607ff.; see Baron-Cohen 2000:1252 for brain-imaging evidence).

■ Languages can switch midstream in interaction ("code-switching"), leaving the interactional framework undisturbed, evidence that interaction structure is independent of the "coded" signal systems of language (Muysken 2000).

■ Ethnographic reports on interaction style rarely question the applicability of the fundamentals. Where they do, as in Basso's (1970) account of massively delayed greetings in Apache, or Albert (1972) on turn taking according to rank in Burundi, or Reisman (1974) on "contrapuntal conversation" in the West Indies, there is reason to believe they are describing something other than the unmarked conversational norm (Sidnell 2001). What the ethnographic reports nevertheless do make a good case for is cultural shaping of all the modalities of interaction, from spacing, posture, and gesture to linguistic form.

■ The small amount of work that has been done on the structure of conversation cross-linguistically and cross-culturally (on Thai, Japanese, Korean, Mandarin, etc.) shows remarkable convergence in many details, supporting the idea of a shared universal framework for verbal interaction (see, e.g., Clancy et al. 1996; Hayashi 2003; Moerman 1989).

■ Humans look different from other primates in the amount of time and effort invested in interaction—it would be interesting to see them in a zoo. We don't actually have good measures of this. Dunbar (1997:116) reports a study of a New Guinea tribe (the Kapanora) whose males spend 30 percent (the women slightly less) of daylight time socializing (or gossiping),[3] compared with 20 percent for gelada baboons (doing grooming), but my suspicion is that such figures hugely underestimate the amount of human social interaction during the business of the day, not to mention the entertainments of the night. Allowing for differences caused by population density, age and gender, and subsistence mode (fishermen may spend the

day alone), and Hymes's (1972) notes about cultural differences in volubility,[4] my guess is that humans on average spend somewhere between 30 percent and 70 percent of waking hours in social interaction, whether at work or play.[5]

I hope this sort of rough and ready list is enough to give the proposal prima facie plausibility—the proposal that, from an ethological point of view, humans have a distinctive, pan-specific pattern of interaction with conspecifics, marked by (1) intensity and duration, (2) specific structural properties, and (3) those properties separable from the language with which it is normally conducted.

Scholars from some disciplines may be puzzled by the absence of *language* from this catalogue of evidence for a human interactional specialization. We are (along with the song birds) a distinctively chattering species. The reason for the demotion of language is that so much attention has been given to it that we have been damagingly distracted from the interactional underpinnings that make it possible. Students of language usage have tried to remedy this, from Grice 1975 to Sperber and Wilson 1995 to my own earlier self—it is quite clear to us that "Language didn't make interactional intelligence possible, it is interactional intelligence that made language possible as a means of communication" (Levinson 1995:232). So language is the explicandum, not the explicans—humans did not evolve language, then get involved in a special kind of social life, it was just the reverse. For language must have evolved *for* something for which there was already a need—that is, for communication in interaction.

Finally, there is another striking kind of evidence for the independence of interaction principles from the specifics of language and culture. Around the world, children are born deaf to hearing parents, who sometimes raise their children without access to a conventional sign language. What emerges is called "home sign," an expressive signing system invented by the child to make himself or herself understood, and that is reciprocated by other means (see Goldin-Meadow this volume). In societies without institutional education of the deaf or a sizeable deaf community, such "home-sign" systems can remain the only communication system for deaf adults. I have investigated a couple of such cases on Rossel Island, a remote island community in Papua New Guinea. Take the case of Kpémuwó, about twenty-eight years old, born in a village where he is the only deaf person, which is three hours walk away from any other deaf people. One day he came to me when I was alone and proceeded to sign. To my intense surprise, I thought I could

understand quite a bit of what he was "saying" although we shared no language, little culture and just a bit of background knowledge. He seemed to be communicating, by means of pointings and iconic gestures, about a woman who was dying of cancer in the neighborhood: There was a lot of detail about the course of her disease, her futile trip to the mainland for treatment, the visits of her daughters, and so forth. Then, when my hosts returned, I got them to "translate" as best they could Kpémuwó's message—according to them, much of what I had inferred was correct. They obtained much further detailed explication, for example about the cause of the impeding death, caused by the antisorcery god Nkaa, depicted by mime as his eagle avatar (see Fig. 1.1), and hence Kpémuwó's reluctance to help the family of the convicted sorcerer.[6]

A moment's reflection will reveal the depth of the mystery here. How is it possible for two people who share no language and little cultural background (myself and Kpémuwó) to communicate at all? For Quine's "radical translation" to be possible after all, despite his scruples, there has to be some powerful meaning-making machinery that we all share. This depends, I claim, on a peculiar ability to match communicative intentions within an interactional framework. Kpémuwó and I got as far as we did because first he signed in such a way as to make his intentions maximally clear to me, and then I gestured my understanding of what

The anti-sorcery God Nkaa in his eagle avatar form
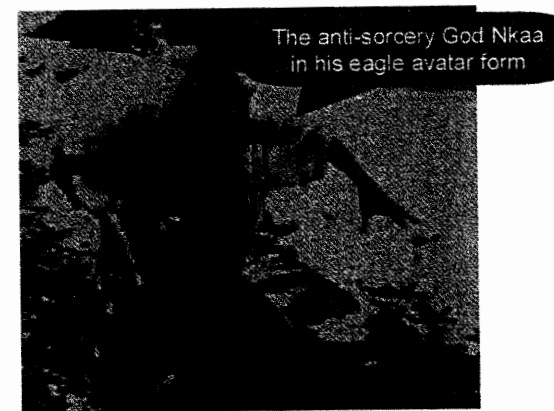
Figure 1.1. Kpémuwó, deaf home signer on Rossel Island, inventing a way to communicate about abstract ideas concerning sorcery.

he signed, and then he in response attempted to correct or narrow my interpretation, until step by step we converged on an understanding. Intention recognition and the mechanics of turn taking are deeply interlocked. The focus of this chapter is on what exactly Kpémuwó and I share that makes it possible for us to communicate, when we share so little other background in conventions of culture and communication.

The Idea of a Core "Interaction Engine": What the Output Shows

The idea in a nutshell is that humans are natively endowed with a set of cognitive abilities and behavioral dispositions that synergistically work together to endow human face-to-face interaction with certain special qualities. I call these elements collectively the human *interaction engine* (which is meant to suggest both dedicated mental machinery and motive power, i.e., both "savvy" and "oomph"). Right away, I should underline this is not a proposal for a "social cognition module," "a culture acquisition device," "cognitive culture system" or an "interaction gene" or anything of that simple-minded sort (see, e.g., Jackendoff 1992; Pinker 1997; Talmy 2000:373ff.). Those accounts assume that the kind of approach taken to the "language module" or "language instinct" can be copied across into a "social–cultural module," and I am arguing nothing of the kind. What I am entertaining is that there are underlying universal properties of human interaction that can be thought of as having a cognitive-and-ethological foundation. Evolution is *"bricolage"* (to use Lévi-Strauss's term), seizing what is at hand in the organism's phenotype to construct an often ramshackle but adaptive system. So an "interaction engine" could be constructed of scraps of motivational tendencies, temporal sensitivities (reaction contingencies), semicooperative instincts, ancient ethological facial displays, the capacity to analyze other's actions through mental simulation, and so forth. The model is a Jean Tinguely kinetic sculpture built of bric-a-brac, not a Fodorean mental module (Fodor 1983), let alone a Chomskyan point mutation (Bickerton 1998; Hauser et al. 2002).

Whatever your doubts, just entertain the idea for a moment (I turn to the crucial question of cross-cultural variability in the next section). Before we ask "What exactly are the elements of the interaction engine?" we need to ask what it needs to account for, that is, what the crucial properties of human interaction are. From the output, we can guess

at the properties of the machine. Here are some obvious properties of the output:[7]

(1) Responses are to actions or intentions, not to behaviors (unlike, e.g., the defensive reaction of a snake to someone who passes too close by). That is, the interpretation of others' behavior is a precondition for interaction. Interpretation involves mapping intentions or goals onto behavior, to yield component *actions*, bundles of behavior and mental instigations (a cough can be just a cough—a reflex—or an intended signal). This parsing of the other's behavior stream clearly presupposes some kind of simulation of the other's mental world.

(2) In interaction, a simulation of the other's simulation of oneself is also involved. This is shown most clearly by the fact that actions are generated taking into account that they will be interpreted by a specific other—that is, they exhibit *recipient design*. So I call my neighbor "Dick" only if I think you will recognize who I mean under that appellation (see Clark et al. 1983; Clark and Wilkes-Gibbs 1986; Schegloff 1972a, 1995). This implies that the interpretation, based as in (1) on actions or intentions, can make use of a further principle: The action to be interpreted can be presumed to have been *designed to be transparent to this particular recipient*.

(3) Although human interaction is dominated by the use of language, language does not actually code the crucial actions being performed—these are nearly always inferred, or indirectly conveyed (Levinson 1983:289–94, 2000; Sperber and Wilson 1995). In addition, "nonce signals" are easily devised, and in the case of "home sign" can even constitute the basis for an individual's main communication system. This implies that the fundamental signaling mechanism is independent of language—language just enormously amplifies its potential.

(4) Interaction is by and large cooperative. This is not a Panglossian claim that we all get on with one another. It is, rather, the claim that there is some level, not necessarily at the level of ulterior motivation, at which interactants intend their actions (a) to be interpretable (the underlying intentions to be recoverable), and (b) to contribute to some larger joint undertaking (having a conversation, making a hut, even having a quarrel!).

(5) Interaction is characterized by action chains and sequences (Schegloff in press, this volume) governed not by rule but by expectation. Thus, there is an assumption that a question expects an answer, but there is no rule that a question must be followed by an answer: "When are you going?" → "Where?" is as well-formed a sequence as "When are

you going?" → "Ten o'clock."[8] The outcome of a momentary interaction is something none of the parties can plan in advance—it is a contingent product. That is why there is no such thing as a formal grammar of discourse.

(6) Interaction is characterized by the reciprocity of roles (e.g., speaker–addressee, giver–taker), and typically by an alternation of roles over time, yielding a turn-taking structure (Sacks et al. 1974).

(7) Interaction takes place within a (constantly modulating) participation structure (specifying who is participating, and in what role), which in turn presumes ratified mutual access (Goffman 1979; Goodwin 1981). We can be copresent on a bus, but not be in such a state of incipient interaction—often rights of mutual access have to be negotiated (e.g., by greetings—see Duranti 1997).

(8) Interaction is characterized by expectation of close timing—an action produced in an interactive context (say a hand wave) sets up an expectation for an immediate response.

(9) Face-to-face interaction is characterized by multimodal signal streams—visual, auditory, and haptic at the receiving end, and kinesic, vocal, and motor at the producing end. These streams present a "binding problem"—requiring linking of elements which belong to one another across time and modality (e.g., a gesture may illustrate words that come later, a hand grasp may go with the following greeting).

(10) Interaction appears to have detailed universal properties, even if little cross-cultural work has actually been done to establish this. What we do know is that for a wide range of features, from turn taking, adjacency pairs (as in question–answer sequences), greetings, and repairs of interactional hitches and misunderstandings, the languages and cultural systems that have been studied reflect very similar, in some cases eerily similar, subsystems.

This list may seem too self-evident and bland to yield any far-reaching conclusions. But there is a lot more to say under each rubric. Let us consider (4), the cooperative nature of most human interaction (at least in the limited sense indicated), in a bit more detail, because of the crucial, and puzzling, role it plays in evolutionary theory (it is just very hard to see how cooperation could ever evolve under natural selection: see Hammerstein 1996; Boyd and Richerson this volume). There seem to be detailed properties of interaction that reflect cooperation, and that contrast with the properties of agonistic interaction. For example, the kind of intended transparency noted in (2) above derives ultimately from (4) cooperation: in antagonistic interaction (as in predator–prey

relations), intentions should be hidden, as opaque as possible (even copresence should be disguised, of course!).

Or consider this: cooperation seems to make possible the specific properties of (5), action chains. In antagonistic interaction, as when tiger chases antelope, we can see interaction chains of the kind: antelope veers right, tiger veers right, antelope wheels left, tiger wheels left, and so forth
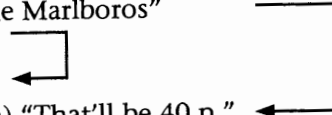
<1>   A1 → B1 → A2 → B2

that is, long chains of immediate responses. What we do not seem to find is anything like the embedded structures typical of human interaction:

<2>   A1 B2 → A2 B1

An example would be the following, in which B's response to A's first action is deferred until clarification has been achieved:

<3>   A: "I'd like some Marlboros"
       B: "10 or 20?"
       A: "20 please"
       B: (passes them) "That'll be 40 p."

The temporary shelving of one interactional task to solve another that is a precondition to it seems to presuppose cooperation. The embedded structure in <2> has formal properties that are quite different from the response chain in <1>: The simple response chain in <1> can be generated by a Markov process, whereas <2> requires something with a push-down stack like a Phrase-Structure Grammar.[9] When it comes to parsing or comprehending a behavior sequence of the kind in <2> as opposed to that in <1>, quite different procedures have to come into play—now a response can be to an action way back in the behavior stream. If this generalizes, then we have a formal test for cooperative structures of interaction—they have long-distance dependencies of this type.

Consider another empirical finding with a bearing on the underlying cooperation in interaction. Conversation analysts have established that after a question, a request, offer, or the like, where a response is

immediately relevant, the response options are not equal but ranked. Responses that are in the expected direction are immediate and brief, responses that are in the opposite direction are typically delayed, marked with hesitations and particles like *well*, and accompanied by explanations. Thus, the absence of an immediate response after the following indirect request apparently indicates quite clearly to the requester that his request will be declined:

<4>   C: "So I was wondering would you be in your office on
            Monday (.) by any chance
            (2 second silence) Probably not." [Levinson 1983:320]

Many details of this kind of asymmetry between what are called "preferred responses" and "dispreferred responses" show that the organization of conversation biases actions in the preferred direction—the system is set up so that it is just easier to comply with requests or accept invitations than to decline them! In short, the system is biased toward cooperation.[10]

Ingredients for an "Interaction Engine"

So far, we have seen that an interaction engine has to predict at least those features of interaction listed above. Now we can ask: what kind of a "machine" could produce those properties, that is, what does the human interactant have to be endowed with to generate such behavior? Let's key the points to our numbered properties above:

(1') To get property (1)—responses are to intentions not behaviors— we need a "Theory of Mind" (ToM). That is, any being capable of attributing goals and intentions to other actors must attribute a mind to the other actor (hence have a folk theory of mind, or ToM, in the broad sense explored by Astington this volume). ToM has typically been operationalized as relativized belief attribution, for example as attributing to Sam the false belief that p (Leslie 2000). But here instead the heart of the matter is *intention attribution*: given the observed behavior, the interaction engine must be able to infer likely goals that would have motivated the behavior.[11] Elsewhere, I have pointed out that this is a highly intractable computational problem, because it amounts to inferring premises from conclusions, which cannot be done by any logical engine (Levinson 1995:230ff.). It could perhaps be done on statistical grounds, using some low-level semiautomatic simulation as in the theory of "mirror neurons."[12] That might account

for a simple class of interpretable actions, like you raising your fork to your mouth. But it would never account for the meaningful cough, or the ironic bow—actions whose interpretations are not in line with the statistical associations. The solution to those must lie in having powerful heuristics. But what exactly? This is where point (2) comes in, the ability not only to simulate the other's point of view but also to imagine what he or she thinks your point of view is.

(2') The simulation of the other's simulation of oneself may seem something more likely to occur in deception than in cooperation.[13] But it is crucial to cooperative interaction. It was Schelling (1960) who demonstrated the empirical power of this heuristic to solve coordination problems implicitly: for example, offer two separated subjects $1,000 if they can both think of the same number without communicating— they can beat the odds (they are likely to assume each will find 1,000 the salient solution). Or ask them to each go to where they think the other will go in a crowded department store (they may fixate on the "lost and found"). Exactly how it works has been much discussed, but clearly it involves a *special kind of reflexive thinking*: thinking what you would be thinking I would be thinking when I did the action. This coordination ability presupposes the notion of *mutual knowledge* (or *common ground*)—the things that I know you know, you know I know, and I know you know I know. But it also involves a notion of *mutual salience*—what leaps out of the common ground as a solution likely to independently catch our joint attention (the number 1,000 or the lost and found). This is what is involved in *recipient design*, the choice of just that phrase that will allow you to find the unique thing I am referring to, when it could be referred to in 1,000 myriad ways, none of them uniquely referring (Clark et al. 1983). A nice example of this is the use of phrases like "The what do you call it," which seem typically used where the speaker estimates the addressee can guess what the speaker has in mind (Enfield 2003). These are mental coordinations, meetings of the mind, in what I shall call the "Schelling mirror world."

(3') Our property (3) was a fundamental underlying signaling system independent of language or conventional code. This is provided by Grice's (1957) theory of meaning, which holds that a signaler S communicates z by behavior B if S intends to cause a recipient R to think z, just by getting R to recognize that intention. In other words, a communicative intention is one that achieves its goal as soon as it is recognized (the action B has no other instrumental efficacy). One way of thinking about this is: S tosses behavior B into the Schelling-mirror world, implying "I bet you can figure out why I did this, just

by knowing that I know you can."[14] Grice's theory gives us an account both of how we can communicate without conventional signals at all (as in *First Contact* or when I met Kpémuwó, the deaf man without a language), and of how we can communicate something distinct from what the conventional signals actually mean (as in irony, metaphor, hints, etc.).

This is how we can understand the meaningful cough, or the ironical bow, in which statistical inference will only allow the attribution of the reflexive cough and the genuine bow. And that is why intention attribution in interaction is altogether a different thing than intention attribution outside interaction. This is the difference between meaning attribution in the Gricean sense and mere action interpretation by an observer. Compare: I appear to smooth down my hair—I could be making the action to smooth down my hair, or I could be signaling "Your hair is standing up." The same behavior has distinct interpretations in (third-person) action-interpretation *simpliciter*, and interaction interpretation.

(4′) Cooperative interaction—our property (4)—differs from antagonistic interaction in precisely the same way: antagonistic interaction only requires mere intention attribution (you'd better believe that however quietly the tiger sneaks up on you, he's out to eat you!),[15] cooperative interaction requires the much more heady *reflexive thinking*: if we are going to carry out a joint action, say building something together, each contributive action has to be so designed that the other can see, just by how it is done, that it is intended to achieve the contributing role it is meant to play (Clark 1996:191ff.).[16] That is another reason, in case there were not enough, why cooperative social systems occupy a remote corner of evolutionarily possible design space (Dennett 1995)—you have to have minds capable of simulating other minds simulating your own.

(5′) What accounts for the fact (our property 5) that interaction is (a) composed of action sequences, and (b) governed not by rules but only by expectations? In principle, actions in interaction could be simultaneous, if complementary, as in duets (see Clark 1996). But they are typically chained, one after the other. It might be thought that, for communication anyway, simultaneous broadcast would mask the message, but that does not deter the cicadas, and it does not explain the human case either, because we can listen and speak at the same time, as in simultaneous translation. No, there must be a reason for the alternation. One fundamental motivation is that, given that what I say has been designed for you to be able to see what I mean, it would be

a good idea to see whether my design was actually as good as I hoped it was, which your response will make clear (Sacks et al. 1974). At the birth of cognitive science, Miller et al. (1960) suggested that the Test–Operate–Test–Exit (TOTE) unit should replace stimulus–response as the basic theoretical unit of human behavior: we test to see if the intended goal was achieved, if not, operate on it and try again. In cooperative interaction, the only way to test is to see what the other person made of our actions. This is part of the motivation for taking turns, and it motivates too the priority accorded in interaction to correction and repair sequences—that is why "When are you going?" → "Where?" is "well formed," or more properly, interpretable.

Linguists and anthropologists had hoped that there might be rules of conversational sequences, like rules of grammar, but the search was fundamentally misguided (Levinson 1983:286–94).[17] There are indeed templates that specify sequences of actions, but these do not have the status of sequential rules (a few rituals excepted). The initial actions in these templates introduce a mental scenario as it were—consider "Are you doing anything tonight?" which introduces the frame Preinvitation → Go ahead → Invitation → Acceptance as in:[18]

A: "Are you doing anything tonight?"
B: "No, why?"
A: "I was wondering if you'd like to catch a movie."
B: "Sure, what's on?"

The whole sequence can gracefully abort at the second turn, without the invitation ever surfacing—a sequential template is a mental entity, around which actions will be directed and interpreted, however the actual sequence transpires.

Which brings us back to Schelling, who pointed out that we can tacitly coordinate by each thinking what the other would think (so, e.g., when inadvertently separated in a department store, we both go to the last place we saw each other). Sequential templates hang in this mirror world of simulated mental spaces, and the ability to traverse those spaces was the theme of (2′). In sum, TOTE gives us sequences, and the Schelling mirror world gives us guiding expectations instead of rules.

(6′) The sequential taking of turns (our property 6) may be partially motivated by TOTE, but turn taking itself does not necessarily imply alternation of roles in a single joint activity (compare taking turns at the gas pump). But some forms of cooperation at least require sharing bites of the same cherry. Sharing a resource is an elemental symbol of sociality,

as in commensality. Informal human interaction is characterized by a conversational mode of exchange, in which the erstwhile speaker becomes a listener, and the erstwhile listener becomes a speaker, the valued commodity apparently being speaking while attentive others hold their tongues. This alternation of roles seems to be universally built into the deictic system of languages ("I" refers to the current speaker, "you" to the current addressee, and my "I" becomes your "you"). Many human societies have asymmetrical assignments of roles and elaborate divisions of labor, but in all of them informal interaction seems to be built on the alternation of conversational roles. Given that human language processing is obligate and automatic (hearing you speak English, I automatically comprehend even if I would rather not), the alternation of listening roles implies an obligatory inhabiting of others' mental worlds. So it seems that cooperative sharing of the communicational resource guarantees our mutual sharing of the Schelling mirror world.

(7') Interaction is organized around a participation structure (our property 7). Part of this is a byproduct of turn taking: in a two-party conversation, the alternation of speaker–addressee roles may exhaust the roles involved. But clearly this is not so when we have three or more participants, then we can have speaker, addressee, and nonaddressed participant. And in addition to passive participants there can be bystanders, nonparticipants with access to the interaction. Much finer discriminations of roles are also possible (Goffman 1979; Goodwin 1981; Levinson 1988). Interaction always presupposes a *participation structure*, which itself presupposes a distinction between being copresent but not in interaction versus copresent and participating. This distinction is precisely what motivates "access rituals" like greetings (Duranti 1997). And this distinction maps once again onto our two kinds of action–interpretation: (a) nonparticipants are engaged in action–interpretation simpliciter, without directly being able to benefit from recipient design, (b) full participants can presuppose that each inhabits the shared Schelling mirror world, allowing each to assume that the other has constructed his or her actions to be interpretable to the intended participants. Thus, participation structure seems to be a property of human interaction that emerges from a number of other properties—turn taking and the mental simulations behind recipient design amongst them.

(8') The close-timing characteristics of human interaction (property 8) may be partially attributed to some independent ethological source, some kind of mental metabolism as it were.[19] But they partly follow from the design-and-test characteristic mentioned in (5'): a response

that indicates how the prior action was understood, needs to be adjacent to that action—and given the free turn competition, the only way to ensure that, is to be immediately next (otherwise another participant may take the next turn, or the prior speaker resume speaking). Hence, the timing properties of human interaction can be partly attributed to the turn-taking properties discussed in (5') and (6').

(9') Humans are not unique in using multimodal signal streams in interaction (our property 9): many animal displays have these characteristics, and since Darwin (1872) there have been many attempts to understand the evolutionary and ethological background. Simultaneous display of bared teeth, flared nostrils, and a growl can serve to signal aggression. But in the human case, at least, the whole display is not just a gestalt: the multimodal behavior streams have to interlock at many minute borders and boundaries. There is no doubt that the digital nature of language is partly responsible for this microstructuring of the signal stream. Yet careful inspection of video records shows that synchrony alone will not do the trick of hooking up the bits in the different signal streams: gestures, facial expressions, nods, and the like can come earlier or later than the words they go with. There seems to be a significant "binding problem" in hooking up the signals that go together. If temporal binding is not sufficient, what will do the trick? Suppose I bow low to the Dean and then wink at you—to see if and how the two signals link requires an analysis of what might have driven each behavior, and how those two intentions might be put together. It is, if one likes, a problem in meaning composition, or goal analysis—the mental reconstruction of communicative intentions expressed through clues which are designed to be just sufficient. Human multimodal communication is as much artifice as ethology, and the capacities that drive it will crucially include the mental simulation of the other (2' above) and the reconstruction of motives or intentions (3' above).

(10') What could be responsible for the apparent universals of interaction, like the turn-taking and repair systems of casual conversation? At least some of the properties seem to follow from, or be motivated by, features we have already considered—thus, turn taking may partly derive from the cooperative sharing of a common resource, and the need to test interpretations that are in a way only guesses in a Schelling mirror world. But it seems likely that many aspects of human interaction (turn taking among them) have a long phylogenetic history. Face and voice recognition, known to implicate specialized circuitry in the brain, are preconditions to any *social* interaction, in the

sense of interaction tailored to specific social persons. Language is also biologically underpinned in many complex ways. The simultaneous activation of gesture, gaze, posture, and paralanguage—the particular channels of human multimodal interaction—all contribute to the distinctive ensemble of interactional signals typical of human ethology, picked up as the flotsam of evolution.[20]

Let me summarize. A review of properties of human social interaction suggests that the core interaction engine consists of a bunch of ingredients, but crucially:

- *Attribution of intention*, or "mind-reading" in a broad sense, is a crucial precondition (call it level 1), but is itself nowhere near the abilities needed to generate human interaction. Here, some automatic link from action–perception to the action–production system, as with mirror neurons, together with statistical induction (as Byrne this volume implies) may be enough to map behaviors onto intentions or goals.
- A crucial additional level is the ability to enter Schelling mirror worlds, to do the mental computations that allow us to simulate the other simulating us. Here, we have the ingredient of *mutual salience for us right now* (reliant on common ground). This allows mental coordination without communication. Now we can beat the statistical odds hands down. The distinction between level 1 above and this second level correlates with antagonistic versus cooperative interaction: level 2 is not necessary for antagonistic interaction (although it may be ruthlessly exploited in it), but it is necessary for cooperative interaction.
- A third crucial level is having *Gricean intentions*, intentions that drive behaviors whose sole function is to have an effect by virtue of having their intentions recognized. It is this level 3 that makes high-level communication possible, and on its foundation language has evolved, and still relies on nearly every occasion of use.
- Woven in and out of this is the cooperative nature of human interaction—there would not be any point of getting into Schelling mirror worlds unless cooperation was a reasonable presumption (once there, sure, Machiavellian exploitation of the system will be a thing to guard against).
- There is a set of empirically observable practices—turn taking, sequence templates and repair among them—which look universal and are only partly derivable from other features. These form part of a raft of ethological proclivities which help to account for the species-specific character of human multimodal communication

Note that this analysis, if correct, has some utility for the comparative study of interaction in phylogeny and ontogeny, for it predicts an ordered series of steps toward human interaction, from level 1–3, while noting that ethological patterns like turn taking may develop earlier and independently.

Culture and the "Interaction Engine"

Many sociocultural anthropologists may react with hostility to the ideas so far discussed—the direction of argument may seem to belittle the role of the cultural construction of social life, like so many ideas in sociobiology, human ethology, or evolutionary psychology. But that would be the wrong reaction. We are trying to probe what underlies that cultural construction, what makes cultures get off the ground so to speak—what makes them learnable, and what provides the framework within which the cultural can do its work. The positive reward for speculating about common human potentialities is that we may understand much better what generates the diversity itself.

Cultural Variation

Interaction is shot through and through with culture. It had better be, because it is the vehicle of culture—without it, there would not be any. Even though culture conditions and shapes private acts—the way we urinate or defecate, for example, or even the way we walk—it is through public, and especially interactive, acts that culture propagates itself. And every anthropologist, indeed every traveler, has been impressed with differences in interactional mores. Just to mention a few of my own observations, consider:

(1) In rural Tamilnadu, in a typical village of 18 castes, who can interact with whom, and in what ways, is elaborately specified in a mental 17×17 matrix (Levinson 1982). One indelible memory is of a high-caste foreman arriving on bicycle at a building site, engendering the total cessation of works as all the low-caste workers scramble down the scaffolding so that they can receive instructions while not having their heads higher than their caste better.

(2) In Cape York, the aboriginal speakers of Guugu Yimithirr incorporate gestures into their verbal interaction in a much more fundamental way than Europeans do. For example, a negative gesture preceding a positive assertion signals a negative proposition, or the subject and object of a verb may be omitted but indicated by gesture. The great majority

of gestures are intended to have directional veracity—no mere hand waving here (see Levinson 2003).

(3) In Chiapas, Mexico, Tzeltal-speaking Tenejapans are peasants who maintain a decorum appropriate to a royal court: Long and elaborate greeting sequences specify whether the intruder is merely passing by (and if so in the same or different direction as the intruded, or past the intruded's home base) or arriving to visit (Stross 1967). Once begun, interaction is properly conducted sitting side by side with the minimum of mutual gaze, each assertion being partially repeated by the recipient, with long sequences of the kind: "I've come to visit you" → "You've come to visit perhaps" → "I have come" → "You have indeed" → "Indeed I have." ... (Brown 1998; Levinson and Brown 2005).

(4) On Rossel Island, Papua New Guinea, interaction is typically dyadic, squatting eyeball to eyeball, with sustained mutual gaze, and incorporating many facial displays, and eye-pointings. Fast, informal, with much mutual touching, two big bankers of shell-money can conduct important business for the whole island with a nod and a wink, making a striking contrast to the apparent Tenejapan formality of interaction over matters much more trivial (Levinson and Brown 2005).

These observations, and a thousand like them, raise the question: What sense does it make to talk about a core interaction engine as if it was a universal property of mankind, given all this rich texture of cultural diversity?

The answer is that the interaction engine is not to be understood as an invariant, a fixed machine with a fixed output, but as a set of principles that can interdigitate with local principles, to generate different local flavors. Let me outline just one example of the kind of interplay between the universal and the culturally particular I have in mind (the details appear in Levinson 2005). Sacks and Schegloff (1979) suggested that two principles govern the reference to persons in English conversation: a preference for using a minimal form (e.g., a name), and a preference for using a form (a "recognitional") under which the referent can be recognized by the recipient. Usually these two preferences can be satisfied simultaneously. But sometimes they come apart. For example, if the speaker is unsure whether the recipient will recognize the referent under a single name, he may try it out, marking the "try" with rising intonation—if there is no uptake, a second name may then be introduced, also with a "try" intonation, then a description, and so on. So we get a sequence like this:

<5> A: ... well I was the only one other than the uhm tch **Fords**?,   ← (1)
      Uh **Mrs Holmes Ford**?   ← (2)
      You know uh the **the cellist**?   ← (3)
                    [
   B: Oh yes. She's she's the cellist   ← (4)
   A: Yes well she and ....

                                    [Sacks and Schegloff 1979:19]

At (1) the speaker tries a single name, upgrading at (2) with a second name and a title, and at (3) with a description, whereupon getting acknowledgement of recognition at (4), the speaker proceeds. What the sequence displays is that recognition takes priority, the minimization being successively relaxed till recognition is achieved (common ground established). It shows that a minimal clue to a Schelling solution is tried first.

Very similar sequences can be found in other quite unrelated languages I have worked on, including Guugu Yimithirr, Tzeltal, and Yélî Dnye (the language of Rossel Island). One has to allow for the fact that upgradings might take different forms (e.g., identifying conventions might employ place of origin specifications), and even that different modalities might be involved (e.g., pointings, eye glances at places of origin)—but allowances made, the sequences are eerily familiar. Here is one from Guugu Yimithirr:

<6> B: ngayu nubuun nhaaway waami dyibaalu warra **Milga-mul?**   ← 1
       1s   one   there   found to.South old   ears-without
       "I came across   one fellow there to the South, old 'without ears'?"
                 <points>
       (0.3)   ← 2
    R: aa   ← 3
       "Oh"
       (0.4)   ← 4
    B: **oo Tommy Confen?**   ← 5
       "old Tommy Confen?"
    R: ee   ← 6
       "ah"
    B: nyulu nhamuun bamaal nganhi wangaarmun nhaathi durrginbigu
       gaadariyga bada
       "That fellow saw me, as I was coming down Indian Head"

                                    [Revgest 00:17:01]

At line 1, B tries Tommy Confen's nickname, namely "Without Ears" (he was deaf), with intonational rise on "Ears." Now critically, he has supplemented this reference with an earlier quick pointing gesture to where the Confen household used to be, coinciding with the underlined word -nubuun—but unfortunately R was not looking (see Fig. 1.2 [a]). B therefore has reason to doubt that R has got the reference: he gazes straight at R throughout this sequence until point 5, to assess whether recognition has occurred (see Fig. 1.2 [b]). R's response at point 3 is slightly delayed, and has a form (indicating "news") suggesting that it could be a response to the earlier part of what B said. B therefore tries again at 5 with rising intonation, with both English names of the referent. R responds positively, with mutual gaze, and B then turns away and resumes the story.

That suggests that cross-culturally there seem to be the same two preferences, they seem to have the same ranking, and when they cannot be satisfied simultaneously, minimization is successively relaxed. Let us take this, on the basis of parallels in four unrelated cultures, as a candidate universal, acknowledging that we would need a lot of further evidence to firm this up.

On Rossel Island, there is an additional wrinkle, a cultural taboo on naming that interacts with these preferences. The taboo specifies that one may not name close in-laws or relatives recently deceased. How does this then interact with the candidate universal preferences? Let us take a look. In the following excerpt, J out of the blue refers to someone as "that (distant) girl," pointing <7> south up over the mountain (utterance [1]).
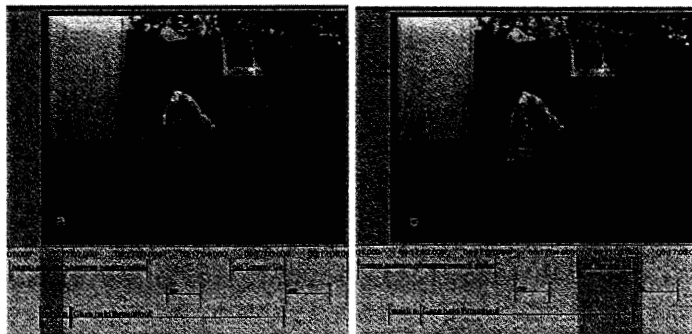
Figure 1.2. Guugu Yimithirr person reference sequence: (a) frame showing unobserved quick point to referent's home base; (b) frame showing mutual gaze at point at which recognition is achieved.

Figure 1.3. Rossel Island name-avoidance sequence: (a) points up on "that thing"; (b) points over mountain on "that girl"; (c) points W on second "that girl"; (d) ditto on "you see," widening eyes; (e) recipient gives eyebrow flash; (f) recipient says "ah" in overlap.

The utterance is "try-marked" with rising intonation, and the gesture is held while looking at R (Fig. 1.3[a], [b]). R does not respond in the gap (2). J upgrades the description in (3), not by adding a verbal description or name but by pointing West while widening his eyes and gazing at R (Fig. 1.3[c], [d]). At this point R responds with an eyebrow flash (a local "yes, continue" marker), followed by a verbal acknowledgment, and recognition achieved, J continues.

<7> J: <u>mu kópu mwo</u>   a pyaa wo, **mu dmâÂdî** ngê?   ← (1)
         that affair over.there happened   that girl   topic
         <—points South over mountain————holds point, looking at R>
         "That thing (pointing) that happened a while ago, that girl?"
         (.)   ← (2)
         mu dmâÂdî ngê? cha w:ee?   ← (3)
         that girl   topic you understand
         <—opens eyes wide, points West>
         "that girl,   you see?"
                          [
    R: (eye-brow flash) <u>éé</u>   ← (4)
                          "right,   yes"
    J: *yi dmâÂdî pi kuu, yed:oo nipi nmî dmâÂdî cha w:ee*   ← (5)
       "that girl is our affair, she's one of ours"

[Rossel Island R02_V4 00:03:27]

The odd thing about the episode is the reference at (1) to a new referent with such a general description ("that girl") with the presumption nevertheless that the referent is recognizable. In holding his gesture, waiting, repeating the description with a new gesture, J is clearly persisting in seeking recognition. Nevertheless, he systematically avoids a name, instead using the same general description but providing two distinct gestural clues, first over the mountain to his own village where the girl was raised, and then West where she has just died (see Fig. 1.3). The recent death (itself only alluded to by "that thing") requires the name avoidance. Thus, R is faced with a Schelling problem: a very general description ("that girl") supplemented with gestural clues, and with the background knowledge that *one reason for not naming a person is their recent death*. The clues evidently prove sufficient, as R claims to have recognized the referent at (4).

Notice how the culturally specific rule (a name taboo) folds into our candidate universal preferences. The speaker goes for recognition. He is blocked from using a name, but uses a brief general description, satisfying minimization, with a gestural clue. When this is not sufficient,

he is again blocked from using a name, and tries an upgrade using a second gestural clue, while claiming with wide eyes (see Fig. 1.3 [d]) and intonation that the addressee can locate the referent in Schelling space. All three preferences are interlocked: do not name, yet go for recognition, while seeking minimal reference. Further cases of name taboo on Rossel show similar patterns: nonverbal upgrades are preferred to verbal ones, as they better satisfy the ban on speaking of the taboo person. Space precludes extensive discussion of this theme, but the point is that the culturally specific does not necessarily eclipse the (candidate) universal procedures—they are woven together to make a coherent local practice.

The identification and naming of persons is, if anything is, a cultural matter, and yet it seems to mesh seamlessly with the universal systematics of interaction. The hypothesis is that the interaction engine will be most recognizable in informal, everyday conversation, which forms the normal matrix for language acquisition and socialization. The ethnography of speaking has long established that when we look at special, ritual or institutional speech events, we find ourselves in the culture-specific territory of séances, ceremonies, investitures, political oratory, and the like (Bauman and Sherzer 1974; Duranti 2001). Even here, though, the interesting suggestion emerging from work in conversation analysis is that specialized speech events are built by *tweaking* the rules and principles governing informal conversation. Thus, the differences between a press conference and a classroom can be partly captured by considering both the similarities (multiple persons, but only two parties, one singular—teacher or press officer) and the differences (questioning assigned to the party with the multiple persons, as in press conferences, or to the singular party, as in classrooms; see Schegloff 1987).

This idea—that the local, cultural specialization is a variation off a universal theme—is potentially powerful, because as we learn more about conversational organization we see that there are relatively few, crucial organizing principles. For example, ringing the changes on different possible systems of turn taking, participation–structure, and action sequences will give us many key aspects of culture-specific speech events. We also see that at a finer level of structure, the modulation of the way in which actions are expressed (e.g., directly vs. indirectly, with or against preference organization) conveys the qualities of social relations (Brown and Levinson 1987). Conversation analysts have therefore sometimes taken a "constructionist" view of social organization (see again Schegloff 1987): you are, as it were, what you say. This does not always accord with the anthropological experience (Levinson 2005): it may work in

New Guinea, but in India you are what you are born. However, viewed as a system of principles that predicts, for each possible manipulation of the systematics, what the consequences will be, it promises to be a powerful tool for understanding cross-cultural variation.

The idea, then, is not that the interaction engine produces cross-cultural uniformity but, rather, that it provides the building blocks for cultural diversity in social interaction. Or in a less crude analogy, it provides the parameters for variation, with default values that account for the surprising commonalities in the patterns of informal interchange across cultures.

One reason that sociocultural anthropologists should be interested in grasping the nature of these parameters is that interactional principles clearly play a central role in higher level social processes. This is entirely transparent in tribal societies, where since Sir Henry Maine (1861) it has been appreciated that larger entities like descent groups act like individuals, contracting marriages and alliances or conducting feuds. Less obviously, politics and diplomacy among modern nation states has much the same character, of a conversation conducted according to the principles of interaction, albeit between representatives of huge agglomerates. We attribute intentions to political maneuvers as if states were individuals, instead of the rambling conglomerates with different factional interests that they really are (Levinson 1995:225).

In short, the analysis of interaction could and should play a major role in our analysis of social institutions and international politics. Humans come natively equipped for interacting with conspecifics. We use this interpretive apparatus for understanding large scale polities of a kind that we have only recently innovated in our evolutionary history, and for which they may be inappropriate. For, however inappropriate, whatever other natural model would we have?

Conclusion

My thesis has been that the notion of a core interaction engine driving human social life makes eminently good sense. There is good prima facie evidence for it, and work in psychology, linguistics, anthropology, sociology and philosophy all point toward it. It is not easy to isolate the critical features of such an ability, because they range from the abstract mental simulations of Schelling mirror worlds, to the concrete problems of binding across multimodal signals, or the processes generating striking cross-cultural parallels across procedures for person reference. But the effort has to be worth it. Progress promises the key to understanding

human evolution, and it offers to shed light on human ontogeny, higher level social processes, and the limitations of a mentality forged in face-to-face contact in the present world of nation states, superhuman agglomerations endowed by us with personal attributes they mostly do not have. It is an effort in which anthropology should have a central role to play.

Notes

1. This chapter takes off from the position paper authored by Nick Enfield and myself, and precirculated in June 2004. Thanks are owed to the participants to the Wenner-Gren symposium in which these ideas were first aired and discussed. I am also grateful for help from Penelope Brown.

2. See Connolly and Anderson 1987. It could have been a chance match of cultural mores, the raw greed of colonial mercantilism happening to meet its match in Melanesian exchange—see Strathern 1992.

3. Dunbar's point is that human verbal interaction replaces primate grooming, and he is therefore especially interested to find that 60 percent of conversation concerns social relationships and person topics (1997:123).

4. Hymes (1972:40) mentions a number of cases in which the ethnographers have noticed the extreme taciturnity of the people—he cites Gardner for example on the Paliyans of South India, who "communicate very little at all times and become silent by the age of 40. Verbal, communicative persons are regarded as abnormal and often as offensive."

5. There are problems quantifying what counts as interaction. Are nonaddressed listeners interacting? Perhaps only if they are ratified participants (see Goffman 1979; Levinson 1988). Is talk essential? No, signs, winks, and nods will do—we are interested in mutual, interlocking sequences of actions (see below), which are not dependent on language. Is a mother rocking a baby "interacting" in the favored sense? Yes, if in response to baby's actions, but no if baby is asleep.

6. My neighbors got further than I did for a number of reasons. First, although Kpémuwó's village is some distance away, he is familiar to them. Second, they shared much more background knowledge of the situations being described. Third, their signing was more perspicuous to Kpémuwó than mine because it made use of conventional elements of the gesture system— the spoken language is accompanied by a rich set of conventional gestures or "emblems."

7. This list, derived from the empirical literature, is not so far removed from the philosophical view derived by H. P. Grice, whose theory of "meaning" (1957) covers points (1) and (2), and whose "maxims of conversation" (1975) cover (3), (4) and (7—"relevance") at least.

8. Conversational analysts have introduced the technical term "conditional relevance" for this expectation (see, e.g., Schegloff 1972b).

9. On the hierarchy of grammars modeling behavior sequences see Partee et al. 1990:433ff.

10. More strictly, the system is asymmetrically structured in such a way that interactants can deploy it to try and extract cooperation (thanks to Tanya Stivers for helping me see this connection between preference and cooperation). See Levinson 1983:332ff. and Schegloff in press for exposition of "preference."

11. False belief tasks are not mastered by normal Western children until almost four years old, but by that age children are experienced interactants. Leslie 1994 suggests that action interpretation begins at around eight months without the notion of propositional attitudes essential to attributions of belief, which begins only at 24 months. Mastery of false belief requires, he argues, a further special kind of inhibition not available for another two years (Leslie 2000:1242).

12. See Rizzolatti and Arbib (1998) on the discovery that specialized neurons fire when the same action is both perceived and executed—suggesting a low-level solution to "reading" other minds. But this correlation is learned, as shown by recent experiments, so there still has to be a higher-level mechanism relating action and perception.

13. In ToM models, this is often called "second-order belief" (what A believes B believes about p: see Baron-Cohen 2000). Here, though, we are actually interested in something that has some of the properties of potentially infinitely nested beliefs: what A believes B believes that A believes … about p. Although that is not psychologically plausible, there are psychologically plausible heuristics that approximate it—see Clark 1996:92ff. for review.

14. Usually this has been thought about the other way around, with Schelling processes embedded in Gricean intentions, rather than the reverse as here suggested.

15. Antagonistic interaction can be Machiavellian, that is, designed to look cooperative but with hidden ulterior motives. In that case it is exploiting cooperative interaction—in a trivial sense, every cooperative interaction can be embedded in a Machiavellian one. The point here is reflexive thinking is not an *essential feature* of antagonistic interaction as it is of cooperative interaction. See following note on the definition of "interaction" here.

16. Why, one might ask, is all this mentalism necessary? Symbiosis after all has two forms, mutualism and parasitism, and both forms, cooperative and antagonistic, can occur without minds. But here I am using "interaction" in a special sense, in terms of sequences of actions, where by definition an action is a pairing of a mental intention or goal and the behavior designed to achieve it.

17. There may be ritual sequences, like greetings and partings, that allow a rule-governed treatment, as in Irvine (1974), but these do not cover the central business transacted in between.

18. Fabricated data for reasons of compression—see Levinson (1983:345); and Schegloff in press.

19. Conversational analysts have noted (of English conversation) that pauses or gaps of between one-tenth to two-tenths of a second—roughly the duration of an unstressed syllable—can often be treated as significant failures to respond. Psycholinguists have tried to link the duration of the segment, the syllable, and the word to the temporal binding properties of the brain—a real temporal metabolism.

20. Earlier attempts to build a science of human ethology (Eibl-Eibesfeld 1989; von Cranach et al. 1979) have largely petered out. Current evolutionary psychology seems headed quite elsewhere, away from the observation of natural human behavior.

References

Albert, E. 1972. Culture patterning of speech behavior in Burundi. In *Directions in sociolinguistics*, edited by J. J. Gumperz and D. Hymes, 72–105. New York: Holt.

Baron-Cohen, S. 2000. The cognitive neuroscience of autism: Evolutionary approaches. In *The new cognitive neurosciences*, edited by M. Gazzaniga, 1249–1257. Cambridge, MA: MIT Press.

Baron-Cohen, S., A. Leslie, and U. Frith. 1985. Does the autistic child have a "theory of mind"? *Cognition* 21:37–46.

Basso, K. 1970. To give up on words: Silence in the western Apache culture. *Southwestern Journal of Anthropology* 26(3):213–230.

Bateson, G. 1955. A theory of play and fantasy. Approaches to the study of human personality. *American Psychiatric Association Report* 2:39–51.

Bauman, R., and J. Sherzer (eds.). 1974. *Explorations in the ethnography of speaking*. Cambridge: Cambridge University Press.

Bickerton, D. 1998. Catastrophic evolution: The case for a single step from protolanguage to full human language. In *Approaches to*

the evolution of language: Social and cognitive bases, edited by J. R. Hurford, M. Studdert-Kennedy, and C. Knight, 341–358. Cambridge: Cambridge University Press.

Brown, P. 1998. Conversational structure and language acquisition: The role of repetition in Tzeltal adult and child speech. *Journal of Linguistic Anthropology* 8(2):197–221.

Brown, P., and S. C. Levinson. 1987. *Politeness*. Cambridge: Cambridge University Press.

Bruner, J. 1976. From communication to language—a psychological perspective. *Cognition* 3:255–287.

Chapple, E. D., and Arensberg, C. M. 1940. Measuring human relations. *Genetic Psychology Monographs* 22:3–147.

Clancy, P., S. Thompson, R. Suzuki, and H. Tao. 1996. *The conversational use of reactive tokens in English, Japanese, and Mandarin. Journal of Pragmatics* 26:355–387.

Clark, H. 1996. *Using language*. Cambridge: Cambridge University Press.

Clark, H., R. Schreuder, and S. Buttick. 1983. Common ground and the understanding of demonstrative reference. *Journal of Verbal Learning and Behavior* 22:245–258.

Clark, H., and D. Wilkes-Gibbs. 1986. Referring as a collaborative process. *Cognition* 22:1–39.

Connolly, B., and R. Anderson. 1987. *First contact*. New York: Viking. [Book accompanying film *First Contact*, Bob Connolly and Robin Anderson, dirs. 54 min. Film Makers Library. New York.]

Darwin, C. 1872. *The Expression of the emotions in man and animals*. London: John Murray.

Dennett, D. 1995. *Darwin's dangerous idea: Evolution and the meanings of life*. London: Penguin.

Dunbar, R. 1997. *Grooming, gossip and the evolution of language*. Harmondsworth: Penguin.

Duranti, A. 1997. Universal and culture-specific properties of greetings. *Journal of Linguistic Anthropology* 7:63–97.

——, (ed.). 2001. *Linguistic anthropology: A reader*. Oxford: Blackwell.

Eibl-Eibesfeld, I. 1989. *Human ethology*. New York: Aldine de Gruyter.

Enfield, N. 2003. The definition of *what-d'you-call-it*. *Journal of Pragmatics* 31:101–117.

Fodor, J. 1983. *Modularity of mind*. Cambridge, MA: MIT Press.

Gardner, H. 1985. *Frames of mind*. New York: Paladin.

Goffman, E. 1979. Footing. *Semiotica* 25:1–29.

Goodwin, C. 1981. *Conversational organization*. New York: Academic Press.

——, (ed.). 2003. *Conversation and brain damage*. Oxford: Oxford University Press.

Goody, E. (ed.). 1995. *Social intelligence and interaction*. Cambridge: Cambridge University Press.

Grice, H. P. 1957. Meaning. *Philosophical Review* 67:377–388.

——. 1975. Logic and conversation. In *Syntax and semantics 3: Speech Acts*, edited by P. Cole and J. Morgan, 41–58. New York: Academic Press.

Hammerstein, P. 1996. The evolution of cooperation within and between generations. In *Interactive minds*, edited by P. Baltes and U. Staudinger, 35–58. Cambridge: Cambridge University Press.

Hauser, M. D, N. Chomsky, and W. T. Fitch. 2002. The faculty of language: What is it, who has it, and how did it evolve? *Science* 298:1569–1579.

Hayashi, M. 2003. *Joint utterance construction in Japanese conversation*. Amsterdam: Benjamins.

Hymes, D. 1972. Models of the interaction of language and social life. In *Directions in sociolinguistics*, edited by J. J. Gumperz and D. Hymes, 35–71. New York: Holt.

Irvine, J. 1974. Strategies of status manipulation in the Wolof greeting. In *Explorations in the ethnography of speaking*, edited by R. Bauman and J. Sherzer, 167–191. Cambridge: Cambridge University Press.

Jackendoff, R. 1992. Is there a faculty of social cognition? In *Languages of the mind*, edited by R. Jackendoff, 69–91. Cambridge, MA: MIT Press.

Kendon, A. 1990. *Conducting interaction*. Cambridge: Cambridge University Press.

Kolb, B., and I. Whishaw. 1990. *Fundamentals of human neuropsychology*. New York: Freeman.

Leslie, A. 1994. ToMM, ToBy, and Agency. In *Mapping the mind: Domain specificity in cognition and culture*, edited by L. Hirschfield and S. Gelman, 119–148. Cambridge: Cambridge University Press.

——. 2000. "Theory of mind" as a mechanism of selective attention. In *The new cognitive neurosciences*, edited by M. Gazzaniga, 1235–1247. Cambridge, MA: MIT Press.

Levinson, S. C. 1982. Caste rank and verbal interaction in Western Tamilnadu. In *Caste Ideology and Interaction*, edited by D. McGilvray, 98–203. Cambridge: Cambridge University Press.

——. 1983. *Pragmatics*. Cambridge: Cambridge University Press.

——. 1988. Putting linguistics on a proper footing. In *Erving Goffman*, edited by P. Drew and A. Wootton, 161–227. Cambridge: Polity Press.

——. 1995. Interactional biases in human thinking. In *Social intelligence and interaction*, edited by E. Goody, 221–260. Cambridge: Cambridge University Press.

——. 2000. *Presumptive Meanings*. Cambridge, MA: MIT Press.

——. 2003. *Space in language and cognition: Explorations in cognitive diversity*. Cambridge: Cambridge University Press.

——. 2005. Manny Schegloff's dangerous idea. *Discourse studies* 7(4–5):431–453.

Levinson, S. C., and P. Brown. 2005. *Comparative response systems*. Paper presented at the 104th Annual Meeting of the American Anthropological Association, Washington, DC, November 30–December 4.

Maine, Sir Henry. 1861. *Ancient law*. London: Murray.

Meltzoff, A., and M. Moore. 1977. Imitation of facial and manual gestures by human neonates. *Science* 198:75–78.

Miller, G., E. Galanter, and K. Pribram. 1960. *Plans and the structure of human behavior*. New York: Holt.

Moerman, M. 1989. *Talking culture: Ethnography and conversation analysis*. Philadelphia: University of Pennsylvania Press.

Muysken, P. 2000. *Bilingual speech: A typology of code-mixing*. Cambridge: Cambridge University Press.

Partee, B., A. ter Meulen, and R. Wall. 1990. *Mathematical methods in linguistics*. Dordrecht, the Netherlands: Kluwer Academic.

Pinker, S. 1997. *How the mind works*. New York: Penguin.

Quine, W. V. 1960. *Word and object*. Cambridge, MA: MIT Press.

Reisman, K. 1974. Contrapuntal conversations in an Antiguan village. In *Explorations in the ethnography of speaking*, edited by R. Bauman and J. Sherzer, 110–124. Cambridge: Cambridge University Press.

Rizzolatti, G., and M. Arbib. 1998. Language within our grasp. *Trends in Neuroscience* 21(5):188–194.

Rochat, P., J. Querido, and T. Striano. 1999. Emerging sensitivity to the timing and structure of protoconversation in early infancy. *Developmental Psychology* 35(4):950–957.

Sacks, H., and E. A. Schegloff. 1979. Two preferences in the organization of reference to persons in conversation and their interaction. In *Everyday language*, edited by G. Psathas, 15–21. New York: Irvington.

Sacks, H., E. A. Schegloff, and G. Jefferson. 1974. A simplest systematics for the organization of turn-taking for conversation. *Language* 50:696–735.

Schegloff, E. A. 1972a. Notes on a conversational practice: Formulating place. In *Language and social context*, edited by P. Giglioli. Pp. 95–135. Harmondsworth: Penguin.

——. 1972b. Sequencing in conversational openings. In *Directions in sociolinguistics*, edited by J. J. Gumperz and D. Hymes, 346–380. New York: Holt.

——. 1987. Between micro and macro: Context and other connections. In *The Micro-Macro Link*, edited by J. Alexander, 207–234. Berkeley: University of California Press.

——. 1996. Some practices for referring to persons in talk-in-interaction: A partial sketch of a systematics. In *Studies in anaphora*, edited by B. Fox, 437–485. Amsterdam: Benjamins.

——. in press. *Sequence organization*. Cambridge: Cambridge University Press.

Schelling, T. 1960. *The strategy of conflict*. Cambridge, MA: MIT Press.

Sidnell, J. 2001. Conversational turn-taking in a Caribbean English creole. *Journal of Paragmatics* 33:1263–1290.

Sperber, D., and D. Wilson. 1995[1986]. *Relevance*. 2nd edition. Oxford: Blackwell.

Strathern, M. 1992. The decomposition of an event. *Cultural Anthropology* 7:244–254.

Striano, T., and M. Tomasello. 2001. Infant development: Physical and social cognition. In *International Encyclopedia of the Social and Behavioral Sciences*. N. J. Smelser and P. Baltes, eds. Pp. 7410–7414. Oxford: Pergamon.

Stross, B. 1967. *Tzeltal greetings*. Language Behavior Research Lab, University of California at Berkeley. [Mimeograph]

Talmy, L. 2000. The cognitive culture system. In *Towards a cognitive semantics*, vol. 2, 373–416. Cambridge, MA: MIT Press.

Trevarthen, C. 1979. Instincts for human understanding and for cultural cooperation: Their development in infancy. In *Human ethology: Claims and limits of a new discipline*, edited by M. von Cranach, K. Foppa, W. Lepenies, and D. Ploog, 530–571. Cambridge: Cambridge University Press.

von Cranach, M., K. Foppa, W. Lepenies, and D. Ploog (eds.). 1979. *Human ethology: Claims and limits of a new discipline*. Cambridge: Cambridge University Press.

Since its inception in 1941, the Wenner-Gren Foundation has convened more than 125 international symposia on pressing issues in anthropology. These symposia affirm the worth of anthropology and its capacity to address the nature of humankind from a wide variety of perspectives. Each symposium brings together participants from around the world, representing different theoretical disciplines and traditions, for a week-long engagement on a specific issue. The Wenner-Gren International Symposium Series was initiated in 2000 to ensure the publication and distribution of the results of the foundation's International Symposium Program.

Prior to this series, some landmark Wenner-Gren volumes include: *Man's Role in Changing the Face of the Earth* (1956), ed. William L. Thomas; *Man the Hunter* (1968), eds Irv DeVore and Richard B. Lee; *Cloth and Human Experience* (1989), eds Jane Schneider and Annette Weiner; and *Tools, Language and Cognition in Human Evolution* (1993), eds Kathleen Gibson and Tim Ingold. Reports on recent symposia and further information can be found on the foundation's website at www.wennergren.org.

The Wenner-Gren Foundation
For Anthropological Research, Inc.

Roots of Human Sociality

Culture, Cognition and Interaction

Edited by

N. J. ENFIELD AND STEPHEN C. LEVINSON

2006 **⊗BERG**

Oxford · New York