

Review

Origins of Human Communication

by Michael Tomasello.

Cambridge, MA, MIT Press, 2008. Pp. 400.

STEVEN GROSS

This bold and wide-ranging book, based on Tomasello's 2006 Jean Nicod Lectures, argues *inter alia* for the following hypotheses:

1. the evolution of human communication required our species-unique capacity for shared intentionality;
2. human communication had its evolutionary origins in gesture, not vocalization; and
3. given the cooperative communicative use of natural gestures backed by shared intentionality, *conventional* communication arose in large part as an emergent feature of human mechanisms of cultural transmission.

The claims are grounded in a wealth of fascinating data, particularly on primate and young child communication and social cognition, much produced by Tomasello's own lab. But there is certainly no dearth of stimulating speculation. Tomasello's story is rich and complex. In what follows, I focus on aspects of the three hypotheses listed above, offering some commentary as I go.

1. Communication and Shared Intentionality

Chimps, our nearest extant relatives, communicate through vocalization and gesture. But they do not communicate in order to be helpful.

... when a whimpering chimpanzee child is searching for her mother, it is almost certain that all of the other chimpanzees in the immediate area know this. But if some nearby female knows where the mother is, she will not tell the searching child, even though she is perfectly capable of extending her arm in a kind of pointing gesture (p. 5).

Chimps simply lack the relevant prosocial motives.

Address for correspondence: Department of Philosophy, Johns Hopkins University, 3400 N. Charles St., Baltimore, MD 21218, USA.

Email: gross.steven@gmail.com

Indeed, according to Tomasello, a failure to factor chimp uncooperativeness into experimental paradigms led him and others mistakenly to conclude a decade ago that chimps lack even rudimentary mind-reading capacities. It turns out that chimps placed in *naturally competitive situations* do indeed seem to understand (in some sense) that others have goals and perceptions on which their actions depend. For example, in competition for food, chimps flexibly adjust their own behaviour according to what a presumed competitor can see and therefore might do.

But as impressive as chimps are, their cognitive skills do not extend, so far as anyone has been able to show, to what is required for 'shared intentionality': they do not engage in collaborative activity or otherwise indicate that they represent themselves and others as thinking and doing things *together*. (Tomasello advances a cognitively lean interpretation of chimp hunting, a proposed counter-example.)

Chimp communication differs markedly in these respects from even simple human communication. Even very young children can point in order to communicate helpfully—to indicate, for instance, where Daddy should look for his glasses. And human communication, again even from an early age, exploits our capacities for establishing joint attention, joint goals, and conceptual common ground. If a 14-month-old and an adult share an exciting experience of an object that is then placed with two other objects, the child will interpret the adult's 'ambiguous' point towards the three objects as indicating the object they experienced together; but she will not so understand the same gesture made by an adult with whom she did not share this experience (pp. 128–9).

Assuming chimps have not lost relevant features our common ancestors possessed (a reasonable assumption insofar as the remarks above generalize to other extant great apes), these differences mark beginning and end points that any account of the evolution of human communication must accommodate. Thus, one of Tomasello's main hypotheses is that the evolution of distinctively human communication required the evolution of prosocial motives and cognitive skills sufficient for shared intentionality. Human communication is a particularly important example of how primate intelligence, conceived in competition (cf. Byrne and Whiten, 1988), was revamped by cooperation. Human ontogeny provides further suggestive evidence for Tomasello's hypothesis: although infants' prosocial motives and physical ability to gesture emerge earlier, their cognitive skills for shared intentionality and their earliest cooperative communication emerge simultaneously around 12 months.

More speculatively, Tomasello claims not just that the motives and skills were evolutionarily necessary for the emergence of cooperative communication, but also that cooperative communication was not necessary for the emergence of these motives and skills, at least in their initial form: indeed, the motives and skills were evolutionarily prior (p. 8). Once collaborative activity involving these motives and skills was in place (Tomasello says little here about *their* origins, though see pp. 193–4), cooperative communication arose to coordinate collaborative activity more efficiently. This occurred, first, in the context of mutualistic collaborative activity, in which participants directly help themselves by helping others. Helpful communication then generalized beyond mutualistic contexts via mechanisms of

indirect reciprocity, whereby one helps oneself by enhancing one's reputation for cooperativeness. Still later, motivations to share emotions and attitudes more generally arose via group-level selection as a means to promote social bonding and expand common ground.

Tomasello's main claim is exceedingly plausible (indeed perhaps tautological under some construals: is communication only cooperative and thus in the relevant sense distinctively human if, unlike chimp hunting, it involves the relevant prosocial motives and cognitive skills?).¹ The obvious strategy in reply would be to argue that the evolution of shared intentionality and collaborative activity require that something like distinctively human communication be in place first or at least coevally.

For example, some argue that, not just the capacity for communication, but more specifically the capacity for *language* is necessary for the capacity to combine representations broadly across modules (Carruthers, 2002; Spelke, 2003). If this is so, and if collaborative activity and/or shared intentionality require a broad capacity for cross-modular representational combination, this would challenge Tomasello's thesis. It is possible, however, that collaborative activity and shared intentionality emerged initially in restricted contexts that required integration only of representations across limited domains—e.g. those involved in the procurement of food. (Note that chimps, on Tomasello's view, presumably already integrate representations relevant to hunting and feeding with representations generated by their capacity for individual intentionality.) Language, if something like Carruthers' or Spelke's view is right, would then serve merely to broaden our already existing capacities for collaborative activity and shared intentionality.

Again, some are sceptical that prelinguistic kids have the capacity for complex meta-representation of the sort apparently implicated in accounts of common ground (e.g. Griffin and Dennett, 2008). Presumably, their scepticism would extend to our prelinguistic ancestors. Tomasello's ascriptions of common ground are perhaps less threatened than others' because he invokes primitive 'we-intentionality,' following Searle (1995). On such a view, it might suffice for joint attention, joint intention, and common ground that the prelinguistic believe that we see X together, intend together to A, know together that P, etc. Ascriptions of such states needn't imply that subjects in addition represent, or have the capacity to represent, that I know that you know that I know *ad infinitum*. (Tomasello sometimes states that common ground *does* require such states, at least implicitly or dispositionally (e.g. pp. 94–5); but elsewhere (pp. 335–7), in what seems his considered view, he suggests that this is a later accomplishment.) Indeed, if we-intentionality involves not a distinct *representation* of togetherness, but rather distinct *attitudes* of seeing-together, knowing-together, etc. (this is Searle's view), then these states are not even higher-order. Some might remain sceptical concerning the ascription of

¹ Cf. p. 15 and also Tomasello's definition of collaborative activity (p. 172) which requires that it involve multiple individuals with *joint goals*.

known-together Gricean communicative intentions, which are at least 5th-order (we know-together that I intend that you know that I intend that you know that P). Phylogenetically, Tomasello hypothesizes that they emerged by the second, indirect reciprocity stage, prior to the emergence of syntax sufficient for ascribing attitudes (see below). Ontogenetically, his position is less clear. Some wording suggests that he would ascribe such intentions to one-year-old infants, while underscoring that they do not achieve an adult-like *understanding* of the states they are in until several years later. But his position might rather be that they only have *proto*-communicative intentions that contain some of the elements of the more mature version (cf. pp. 130–5, 144–5).

2. Gesture First

Tomasello argues that at least the first two stages of cooperative communication's emergence (in mutualistic contexts and to secure reputation)—as well as the first conventionalization of signs (on which, more below)—proceeded in the gestural, not the vocal domain. One of his main grounds is again comparative. Great ape vocalizations are highly genetically constrained, inflexible, and closely tied to specific emotional states. Not so their gestures—at least some of them. These gestures are learned, used for various communicative purposes, and under voluntary control—in particular, deployed or not depending on their intended recipient's attentional state.² They therefore already possess some features found in human communication. Ape-like gestures among our ancestors would seem well-poised for exploitation in early human collaborative activity. Ape-like vocalizations, on the other hand, could only serve this expanding function if they were first brought under voluntary control.

The simpler path does not settle the matter. Of course, at some point, vocalizations in our lineage *were* brought under voluntary control. So, why not earlier rather than later? Tomasello has a second argument for gestural origin—viz. that gesture provides a better medium among humans for referential communication than non-conventional vocalization. This is because pointing *naturally* directs human attention to external targets, whereas vocalizations do not. (Neither does with non-human primates.) Moreover, pantomiming, supported by our natural tendency to infer intention, provides broader opportunities for successfully directing human

² Tomasello distinguishes attention-getters and intention-movements. With the former—ground-slapping, hand-clapping, etc.—an ape, having previously noticed the effects of such behaviours, takes advantage of what naturally draws conspecifics' attention in order to draw attention to itself. The latter—e.g. back-touching in order to get mother to lower her back, arm-raising in order to initiate play—emerge from a process of anticipatory foreshortening. A mother, for instance, comes to anticipate her child's attempts to climb up her back and so lowers herself at the first touch; the child in turn comes to anticipate this and so simply produces the back-touch gesture.

imagination than does vocal mimicry. Tomasello concludes that, since natural communication preceded conventional communication (pp. 58–9), this provides grounds for thinking that early human communication was gestural.

However, while it may be more likely that early human communication was exclusively gestural rather than exclusively vocal, Tomasello's second argument does not preclude a mixed origin. A similar remark applies to Tomasello's thought experiment in support of the claim that it's easier to imagine a path from natural to conventionalized gestures than from unlearned to learned, conventional vocalizations. Tomasello asks the reader to imagine two groups of pre-linguistic human children stranded on an island with their needs somehow arranged for. One group has their mouths bound, the other their hands tied. Which is more likely to succeed in communicating and to establish communicative conventions? But left out of the thought experiment is a group that can use *both* their mouths and hands.³

Tomasello makes various other points regarding the gesture-first hypothesis. I mention just one more. The human version of the FOXP2 gene, believed to be implicated in fine motor control of speech articulators, reached fixation no more than 150,000 years ago. According to Tomasello, this suggests that vocal dominance emerged very late in the evolution of human communication. But the fixation date is consistent with a mixed gestural-vocal origin and even with a long period of vocal dominance prior to the mutation. It is tempting to add that 150,000 years ago counts as late only relative to a question-begging timeline. But here one must recall that Tomasello's topic and thus his timeline encompasses the emergence of cooperative communication generally, not just the emergence of language. In fact, Tomasello hypothesizes that vocal dominance coincided with the (late) emergence of conventional language with something like contemporary syntax.

Suppose human communication did have a gestural origin. Genetic enablers aside, why and how did vocal communication come to dominate? Tomasello cites various possible advantages others have suggested: it frees the hands for simultaneous communication and manual manipulation; it frees the eyes for gathering environmental information; it enables communication over longer distances, in the dark, through dense forests, and around barriers; and so on. To this list of non-exclusive candidates, Tomasello adds the fact that vocalization is arguably more public than gesture, something that might have become advantageous once issues of reputation became relevant to human groups.

Such speculations are exceedingly difficult to evaluate, both owing to the state of the evidence and to the difficulty of modelling the candidates' interactions with other pressures. Moreover, they raise the question: if these factors are so

³ Arguably, young children begin vocalizing for the purpose of cooperative communication around the same time they begin gesturing cooperatively (Clark, 2009, pp. 97–8)—even if, as Tomasello notes (pp. 161–2), they begin *verbalizing* only afterwards. Indeed, the vocalizations and gestures are often produced in tandem, just as shortly thereafter 'many of children's earliest one-word utterances are actually combinations of pointing and language (as well as intonational marking of motive)' (p. 264). But of course one can't read our phylogeny off our ontogeny.

advantageous, why didn't vocalization come to dominate from the start? The most natural answer would posit changes in the fitness landscape, perhaps in part brought about by the emergence of cooperative gestural communication itself. Thus, regarding Tomasello's suggestion, recall that the concern for reputation emerges only in the second stage of his story.

As to how vocal dominance emerged, Tomasello proposes, without much development, that vocalizations first piggy-backed on action-based gestures (pantomimes) as emotional accompaniments and perhaps, once vocalization was brought under voluntary control, as sound effects (e.g. mimicry of animal vocalizations). Insofar as recipients came to associate gesture and sound and see them as redundant, the vocalizations could come to function on their own as the gestures had—and would do so if it were advantageous. Presumably, the use of specific vocalizations would spread by cultural transmission, with biological adaptations perhaps over time facilitating the whole process.

But even if this mechanism is part of the story, it's unclear how it could suffice to explain the emergence of vocal *dominance*. There are only so many distinct vocal expressions of emotion and only so much vocal mimicry of which even modern humans are capable (and only so many nameable things we associate with distinctive sounds). Further mechanisms are needed for the creation and transmission of novel vocal signs. Otherwise, for all that's been argued, our capacity for creative pantomime might outrun our capacity for redundant vocalization. Tomasello might here simply invoke his tentative suggestion that there arose:

... some kind of general insight ... that most of the communicative signs we use have only arbitrary connections to their intended referents and social intentions, and, so, *voilà*, we can if we want make up new arbitrary ones as needed (p. 223).

3. Conventional Communication

A further distinctive feature of contemporary human communication is its exploitation of *conventionality*. We use both basic signs and grammatical constructions that are shared and arbitrary (in particular, non-iconic, unlike pantomimes). Tomasello's third main hypothesis concerns their emergence. He argues that, given the communicative use of natural gestures backed by shared intentionality, *conventional* communication arose in large part as an emergent feature of human mechanisms of cultural transmission.

For holophrastic signs—in particular, for the transition from pantomimes to conventional gestures—Tomasello illustrates his proposed mechanism as follows. A person pantomimes digging for tubers in the direction in which they are normally found. Her cavemates understand that she thereby intends to communicate that they should go dig tubers there. Some of them learn this gesture from her by role-reversal imitation, so that there is now a *shared* communicative device. The gesture

can then lose its *iconicity* if it is misconstrued by ‘some others not familiar with digging, perhaps children’ (p. 223), who take it, say, to indicate leaving generally and then imitatively use it themselves for that purpose.

It’s important that Tomasello has, not one, but *some* others misconstrue the gesture; that they misconstrue it *in the same way*; and that the misconstrual isn’t completely unrelated to the intended content (leaving to dig tubers is a *kind* of leaving). These features raise the likelihood that a future use of the gesture with the new purpose will succeed and thus catch on. The odds in any particular case might remain low, but the account only requires that the mechanism *sometimes* leads in this manner to loss of iconicity. If, however, there were mechanisms that did not involve misconstruals, the odds might be significantly higher *ceteris paribus* (cf. p. 296). Tomasello’s proposed mechanism might then be part of the story, but not the only and perhaps not the main part. Tomasello’s discussion elsewhere indeed suggests such alternative mechanisms. First, consider the replacement of gestures by piggy-backed non-iconic vocalizations, mentioned above. This mechanism does not require that the gesture be one that has itself already lost its iconicity (cf. pp. 234, 237, 241). It might be replied, however, that, for the same reasons we questioned the prevalence of this process above, it’s unclear how many conventional signs could have arisen this way. Second, there might be processes of gradual abstraction, compression, etc. imposed by pressures for simpler production and enabled by common ground and the familiarity bred by frequency (cf. below). If a digging gesture can be construed as a gesture for leaving, as in Tomasello’s proposal, then a gesture for digging so abstracted from more accurately pantomimed digging as to have lost its iconicity for new users (perhaps it becomes just a downward wrist-flick with palm faced upwards) can be learned by them as well.⁴ That said, since semantic shifts of course do occur, there is no reason to doubt that Tomasello’s mechanism is at least *among* the mechanisms involved in conventionalization.

What of grammatical conventions? Tomasello argues that sign-combinations of increasing complexity emerged owing to functional pressures imposed by our three main communicative motives (associated respectively with the three stages of expanding cooperative communication mentioned above): requesting, informing, and sharing emotions and attitudes. Basic requests in mutualistic contexts focus on objects and actions in the here-and-now that can advance shared goals. Often a single gesture (pantomimed drinking, to borrow an example from Tomasello) can suffice, given common ground, for the recipient to discern communicative intention (that the recipient should fill her a glass of water, as the case may be). But in situations with multiple relevant objects and actions, there is pressure to combine gestures in order to provide more information concerning desired events and participants (e.g. one might pantomime drinking and point to a pitcher of

⁴ Tomasello (p. 223) does once refer to the digging gesture as ‘ritualized,’ but this is not given any role in his account of how iconicity is lost. Indeed, Tomasello (pp. 223–4) argues that it’s the loss of iconicity that allows for stylized, abstracted signs, rather than the other way around.

water). With informing, there is pressure for devices that can help identify things that are not present, indicate the relations among participants in events (who did what to whom), and distinguish among speaker's motives (e.g. requesting versus informing). Finally, the motive of expressing and sharing emotions and attitudes, particularly through narrative, introduces pressure for devices that enable relations to be tracked across *multiple* events and participants.

Various devices can answer to these pressures, as witnessed by the multitude found among extant languages. The first, on Tomasello's account:

... were derived from 'natural' principles—that is, ones that all human beings naturally employ based on their general cognitive, social, and motivational propensities, such as 'actor first' or 'topic first' or looking puzzled when asking for information—but the conventionalization process then transformed these into communicatively significant syntactic devices in human cooperative communication (p. 282).

Conventionalization, Tomasello argues, proceeds as follows. Successful combinations are abstracted into shared constructions—structured schema with variable slots for individual signs (cf. Tomasello's (2003) usage-based account of language acquisition). New constructions less directly tied to 'natural' principles then emerge by processes of the sort linguists invoke to explain language *change*—e.g. grammaticalization. Among speakers with common ground, the use of a high frequency shared construction increases the predictability of communicative intention. This, given speakers' tendency to expend the least effort necessary for successful communication, leads speakers to reduce the construction's form. 'He pulled the door and it opened,' for example, becomes the resultative construction 'He pulled the door open.' Further changes introduced by slippage in the process of transmission can then likewise move constructions further away from their 'natural' roots. Language learners attempt to understand both the overall meaning of an utterance and the contributions made by constituents. Tomasello provides an example in which their success with the former involves a functional *reanalysis* regarding the latter:

... if a child hears an adult say 'I'd better go,' she might not hear the *-d* so well and just assume that *better* is a simple modal auxiliary like *must*, as in 'I must go' ... if there are many similar children, at some historical point *better* will indeed become a modal auxiliary like *must* in the English language at large (pp. 304–5).

Note that this case does not involve a misconstrual of overall meaning (cf. above). But nothing precludes such cases; and Tomasello cites the semantic shifts that accompanied the grammaticalization of the future marker 'gonna' (pp. 302–3, albeit while discussing reduction, not reanalysis in transmission).

It will not go unnoticed that no mention has been made of innate language-specific biases.⁵ Tomasello dismisses these in two brisk paragraphs, at least in the case of syntax (pp. 311–3; for more, see Tomasello, 1995, 2003, and 2004). In his view, general cognitive, social, and vocal-auditory constraints, together with shared functional demands, suffice to explain, for example, linguistic universals. Of course, the jury remains out on this contentious topic. So, it's of interest to ask how much of Tomasello's view could be retained should his position on innate language-specific biases prove mistaken. The answer, as far as I can see, is pretty much all of it. Consider the common gradualist claim that innate language-specific biases emerged via some version of the Baldwin Effect (e.g. Waddington, 1975; Pinker and Bloom, 1990; and Jackendoff, 2002). On such views, the rise of conventional communication alters the fitness landscape in a way that creates selective pressure for mechanisms that ensure reliable and perhaps more rapid acquisition of the capacity for conventional communication. Whether such effects could plausibly occur for language, given, for example, how quickly particular languages themselves change, is currently a central topic in computational modelling approaches to language evolution (Briscoe, 2003, 2009; Christiansen and Chater, 2008). But, assuming they can and did, the gradualist-nativist can nicely blend her account with Tomasello's. The processes Tomasello describes would account for the rise of conventional communication—i.e. for what changed the fitness landscape so as to give rise to innate language-specific biases. Those processes would continue to operate, albeit now within the further constraints the biases impose. Of course, Tomasello's story may be consistent with innate language-specific biases while conflicting with more specific proposals that incorporate them.

However this issue and the others mentioned above sort out, we can be grateful to Tomasello for this engaging, highly stimulating book.

*Department of Philosophy
Johns Hopkins University*

References

- Briscoe, T. 2003: Grammatical assimilation. In M. Christiansen and S. Kirby (eds), *Language Evolution*. Oxford: Oxford University Press.
- Briscoe, T. 2009: What can formal or computational models tell us about how (much) language shaped the brain? In D. Bickerton, S. Kirby, and E. Szathmáry (eds), *Biological Foundations and Origins of Syntax*. Cambridge, MA: MIT Press.

⁵ In this context, 'language-specific' means specific to language as opposed to vision, say, or causal reasoning (and as opposed to general—i.e. not specific to any module or faculty). It does not mean specific to a particular language, such as English.

- Byrne, R., and Whiten, A. (eds) 1988: *Machiavellian Intelligence: Social Expertise and the Evolution of Intellect in Monkeys, Apes, and Humans*. Oxford: Oxford University Press.
- Carruthers, P. 2002: The cognitive functions of language. *Behavioral and Brain Sciences*, 25, 657–726.
- Christiansen, M., and Chater, N. 2008: Language as shaped by the brain. *Behavioral and Brain Sciences*, 31, 489–558.
- Clark, E. 2009: *First Language Acquisition*, 2nd edn. Cambridge: Cambridge University Press.
- Griffin, R., and Dennett, D. 2008: What does the study of autism tell us about the craft of folk psychology? In T. Striano and V. Reid (eds), *Social Cognition: Development, Neuroscience, and Autism*. Oxford: Wiley-Blackwell.
- Jackendoff, R. 2002: *Foundations of Language*. Oxford: Oxford University Press.
- Pinker, S., and Bloom, P. 1990. Natural language and natural selection. *Behavioral and Brain Sciences* 13, 707–84.
- Searle, J. 1995: *The Construction of Social Reality*. New York: Free Press.
- Spelke, E. 2003. What makes humans smart? Core knowledge and natural language. In D. Gentner and S. Goldin-Meadow (eds), *Language in Mind*. Cambridge, MA: MIT Press.
- Tomasello, M. 1995: Language is not an instinct. *Cognitive Development*, 10, 131–56.
- Tomasello, M. 2003: *Constructing a Language: A Usage-Based Theory of Language Acquisition*. Cambridge, MA: Harvard University Press.
- Tomasello, M. 2004: What kind of evidence could refute the UG hypothesis? In M. Penke and A. Rosenbach (eds), *What Counts as Evidence in Linguistics: The Case of Innateness*. Amsterdam: John Benjamins.
- Waddington, C. 1975: *The Evolution of an Evolutionist*. Edinburgh: Edinburgh University Press.