3

The Key Is Social Cognition

Michael Tomasello

3.1 Introduction

Surveying human evolution and history, it is difficult to find a good analogy to language. But the closest might be money. Economic activities—in the broad sense of people exchanging goods and services with one another—antedate the invention of money by many millennia, and economic activities do not absolutely require money. But the invention of money as a symbol for exchanges, and its historical development into more complex forms such as paper and electronic money, is clearly responsible for some new forms of economic activity. Certainly, modern economies could not exist as they do without something resembling the monetary symbol systems currently in use.

Let's try another, more cognitive analogy. Basic quantitative skills are possessed by all mammals and even some bird species, and so they assuredly do not rely on written symbols and notations. But when human beings invented written symbols and notations to help them count and calculate, all of a sudden they began to count and calculate in some new and more complex ways. And it is well known that some notation systems enable certain kinds of calculations that others do not. For example, it is basically impossible to imagine doing algebra or calculus (not to mention long division) with Roman numerals; something like Arabic numerals, based on the place value system (and with a zero), is required for modern mathematics.

The way human beings behave and think thus changes when symbols, including linguistic symbols, become involved. Money and mathematics are two good examples, but the analogy to language is not perfect.

Spoken language is more basic than these. In many ways, minted money and Arabic numerals are more like written language than spoken language—and indeed, historically the invention of written symbols for speech, mathematical activities, and economic activities were closely intertwined. Ontogenetically, human beings acquire competence with a spoken language much earlier than with written symbols, and this happens in close concert with their earliest understandings of many aspects of their physical and social worlds.

I have a specific hypothesis about how language transforms cognitive activity during human ontogeny. (Perhaps the hypothesis may be extended to human evolution, but that is another story.) It is based on the conviction that the two main functions of language—for communication and for cognitive representation—are closely interrelated. The key is the nature of uniquely human social cognition and how it enables the learning and use of linguistic symbols for purposes of interpersonal communication beginning in the second year of life. Gradually, these interpersonal instruments are internalized and used intrapersonally, becoming the major representational medium for certain kinds of human cognition. The hypothesis is thus explicitly Vygotskian in spirit, but I am focusing on much younger children than Vygotsky ([1934] 1962, 1978) and, I would argue, on much more fundamental aspects of human cognition.

3.2 Joint Attention as Social Cognition

All primate species cognitively represent the world. They recall where things are located after significant delays, they anticipate impending events, they use spatial detours and shortcuts creatively (cognitive mapping), they categorize novel objects on the basis of perceptual similarity, and they solve novel problems on the basis of mental trial-and-error or "insight" (Tomasello and Call 1997). These activities all involve the ability to cognitively represent the world in the sense that past sensorimotor experiences can be preserved and, in some cases, actively *re*-presented and consulted as a guide for behaviors such as navigation, search, or problem solving.

Before they begin to use language, human infants cognitively represent the world in many of these same kinds of ways, and—importantly—in no others. However, socially and communicatively, the prelinguistic infant is doing things that other primate species are not. Specifically, she is both following into adults' attention to outside objects and events—for example, by following their pointing gestures or imitating their actions on objects—and actively directing their attention to outside objects and events —for example, by pointing and showing objects and events to others. No other species on the planet points to or shows objects and events to its conspecifics in these ways. This is a fact with far-reaching implications for our understanding of language and its relation to human cognition.

Carpenter, Nagell, and Tomasello (1998) followed longitudinally the emergence of nine different "joint attentional" behaviors, along with the emergence of linguistic skills, in infants from 9 to 15 months of age. The joint attentional behaviors included both those in which the infant followed into the adult's behavior or attention (e.g., gaze following, imitation of actions on objects) and those in which the infant directed the adult's attention via such actions as pointing and showing. Language comprehension and production were also assessed. The relevant findings were as follows: (1) All nine of the nonlinguistic joint attentional behaviors emerged in individuals as a group (mostly within a three- to four-month time window). (2) These behaviors emerged in a predictable order: first, behaviors that involved the infant checking that the adult was attending; next, those in which the infant followed adult attention to outside objects and events; and finally, those in which the infant actively directed adult attention to outside objects and events. (3) There was a very strong correlation between infants' ability to engage in joint attentional activities with their mothers and the emergence of language comprehension and production.

Carpenter, Nagell, and Tomasello (1998) argued for two key theoretical points relevant to the current discussion. First, they argued that the reason that all of these joint attentional behaviors emerge together is that they all represent—each in its own way—manifestations of infants' newfound understanding of other persons as intentional agents, like themselves, whose attention to outside objects can be followed into,

directed, and shared. Other primate species do not understand others of their species in this same way, and so they do not engage in this complex of joint attentional activities. Second, Carpenter, Nagell, and Tomasello argued that the reason linguistic skills emerge on the heels of these joint attentional activities—and strongly correlate with them—is that language is itself a form of joint attentional activity. The infant's first comprehension of language is nothing other than her emerging understanding of utterances as indications of other persons' intention that she join them in attending to something. The infant's first production of language is nothing other than her emerging ability to express her own communicative intention that other persons join her in attending to something. Simply said, language is about sharing and directing attention, and so it is no surprise that it emerges along with other joint attentional skills.

It is worth pausing for a moment to note that theories of language development that neglect these joint attentional skills have basically no explanation for why language emerges when it does. Prelinguistic infants, and many animal species, are able to (1) perceive outside entities, (2) perceive discrete vocal sounds, and (3) associate sounds with visual experiences. So why can they not learn a piece of language? The reason is that associating sounds with experiences is not language. It is simply associating sounds with experiences—in much the way that a pet dog associates the sound "Dinner!" with the experience of food. Learning a piece of language requires that the learner understand that the other person is making this sound in order to direct her attention to something on which he, the speaker, is already focused. Theories that do not appreciate the essential, indeed constitutive, role of this kind of social cognition in the process of language acquisition simply cannot explain why most children begin learning language only after their first birthday and not earlier.

And so, the first point is that language is acquired as a joint attentional activity for sharing and directing the attention of other persons. The social-cognitive ability that enables such activities—the understanding of other persons as intentional agents like the self—is specific to human beings and emerges reliably at around 9 to 12 months of age in human ontogeny.

3.3 Organizing Cognition for Purposes of Linguistic Communication

What makes skills of linguistic communication different from other joint attentional skills, of course, is the fact that the language the child is learning antedates her arrival on the scene. Each of the world's 6,000+ languages is a collection of symbols and constructions that some community of people has created over historical time for purposes of sharing and directing the attention of one another. What is so mind-boggling about languages is the unbelievable number of different ways each one of them has accumulated for accomplishing these tasks.

To concretize the discussion, let us focus on the process of establishing a topic by means of what philosophers call an "act of reference"—in this case, to a single individual object in the world. One could imagine a language in which this was done with proper names only. That is, just as we refer to highly familiar people, places, and things with proper names such as Bill, Fluffy, Disneyland, and Big Bertha (a golf club), so we might refer to each object in the world with its own proper name. But presumably this would at some point overtax human memory. Consequently, individuals typically have proper names for only a few hundred or perhaps thousand especially familiar persons, places, and things in the world (a different set for different individuals), and in addition they use many thousands of category labels (common nouns). To use these category labels to direct another person's attention to an individual object (in cases in which its identity is in question), they must then use some further means of specification, either linguistic (e.g., *the X in my room*) or nonlinguistic (e.g., pointing).

It is an astonishing property of these category labels that they are used to refer to objects from many different perspectives depending on both the communicative context and the communicative goals of the speaker. As just a sampling, common nouns embody attentional construals based on such things as (1) granularity-specificity (*thing*, *furniture*, *chair*, *desk chair*), (2) perspective (*coast*, *shore*, *beach*, *vacation lot*), and (3) function (*father*, *lawyer*, *man*, *American*) (Langacker 1991).

Then, once a speaker and a hearer have established shared attention to a particular X, they can make future references to it in the same context with some short, all-purpose, easy-to-use symbol like *it*.

The outcome is that an individual language user looks at a particular tree and, before drawing the attention of her interlocutor to that tree, must decide, based on her assessment of the listener's current knowledge and expectations (and her own communicative goal), whether to use *that tree over there*, *the oak*, *that hundred-year-old oak*, *the tree*, *the bagswing tree*, *that thing in the front yard*, *the ornament*, *the embarrassment*, *that one*, *that*, *it*, or any number of other expressions. In terms of the new information being communicated, when the speaker wants to predicate (focus the hearer's attention on) something about the tree, she must decide if the tree *is in*, *is standing in*, *is growing in*, *was placed in*, or *is flourishing in*—or whatever—the front yard. And in the construction chosen to unite topic and focus, many more specific perspectives arise as the speaker attempts to highlight or downplay specific referents (e.g., *A tree is in the yard* vs. *In the yard is a tree*)—and even more than that when multiple participants are involved, as in *He broke the vase*, *The vase broke*, *It was him that broke the vase*, *It was the vase that got broken*, and on and on. Finally, based on categories of consistent functions in constructions, speakers may actually construe things in ways other than their "normal" ontological class. For example, the speaker can conceptualize objects as actions (as in *He porched the newspaper*), actions as objects (as in *Skiing is fun*), and attributes as objects (as in *Blue is my favorite color*). Ultimately, users of a language become able to conceptualize all kinds of abstract situations in terms of concrete metaphorical construals—for example, *Love is a journey*, *The office is pressing on me*, or *My spirits are high* (Lakoff 1987).

The main point is that decisions to construe a referent or predication in one way rather than another possible way are not made on the basis of the speaker's direct goal with respect to the object or activity involved; rather, they are made on the basis of her goal with respect to the listener's interest and attention to that object or activity. The speaker knows that the listener, as a user of the same language, shares with her the same range of choices for construal; these are in the form of the myriad perspectives that are symbolized in the known, but on this occasion not chosen, symbols and constructions in the shared language. Human linguistic symbols are thus both *intersubjective*—(both users know, and each knows that the other one knows, the range of possibilities) and

perspectival (any one symbol embodies one way, out of these many simultaneously available ways, that a situation may be construed for a given communicative purpose). It is difficult to think of how or why, in the absence of linguistic communication, a child or adult might choose to construct all of the many different perspectives on things that are routinely symbolized in human languages.

3.4 New Forms of Thinking

The specific developmental hypothesis is this. As the young child internalizes a linguistic symbol or construction—as she culturally learns the human perspective embodied in that symbol or construction—she cognitively represents not just the perceptual or motoric aspects of a situation, but also one way, among other ways of which she is also aware, that the current situation may be attentionally construed by "us," the users of the symbol. The intersubjective and perspectival nature of linguistic symbols thus creates a clear break with straightforward perceptual or sensorimotor cognitive representations. It removes them to a very large extent from the perceptual situation at hand, and in ways much more profound than the fact that they can stand for physically absent objects and events (and other simple forms of spatiotemporal displacement). Rather, the intersubjective and perspectival nature of linguistic symbols actually undermines the whole concept of a perceptual situation, by layering on top of it the multitudinous and multifarious perspectives that are communicatively possible for those of us who share a certain set of linguistic symbols.

This is the major sense in which, as I proposed initially, the communicative and cognitive functions of language are inextricably intertwined (see also Tomasello 1999). But it must be emphasized that acquiring skills of linguistic communication does not magically create new cognitive skills and representations out of nothing. Acquiring a language in the first place requires the whole panoply of basic primate cognitive skills of perception, categorization, memory, relational understanding, problem solving, and so on. In addition, it requires the uniquely human form of social cognition—understanding other persons as intentional agents like the self—without which there could be no humanlike forms

of cultural or symbolic activity at all. But then, in my hypothesis, as these skills are used to acquire a historically evolved set of linguistic symbols and constructions, children begin to think in some fundamentally new ways.

This is not the end of the ontogenetic story, of course. We might also go on to somewhat older children and examine how they internalize their entire symbolic and linguistic dialogues with other persons in Vygotskian fashion. In Tomasello 1999, I attempted to show how this "second stage" of internalization leads young children to engage in such qualitatively new forms of cognitive activity as multiple classifications of objects and events, metaphorical construals of abstract situations and events, dialogic thinking, and reflecting on one's own thinking. I also attempted to show that the internalization process is not something mystical, as some seem to believe; rather, it is simply the normal process of cultural learning from others when the activity being learned is one that involves perspective taking. Reflecting on one's own thinking derives from internalization in the special case of an activity in which another person takes a perspective on you and your cognitive activities—as in so-called instructed learning (Tomasello, Kruger, and Ratner 1993).

But all of this is a booster rocket on top of the primary reality that emerges at the end of infancy: the intersubjective and perspectival forms of symbolic communication that children engage in with other persons—and the internalization of these into the kind of flexible and powerful cognitive representations characteristic of, and characteristic only of, the species *Homo sapiens*.

3.5 Linguistic Cognition

We may say that thinking is essentially the activity of operating with signs.
Ludwig Wittgenstein

We have no power of thinking without signs.
Charles Sanders Peirce

Only in terms of gestures as significant symbols is the existence of mind or intelligence possible.
George Herbert Mead

Thought is not merely expressed in words; it comes into existence through them.
Lev Vygotsky

These four thinkers have thought as much about thinking as anyone. What could they possibly mean by saying, essentially, that thinking is only possible with symbols? They presumably do not mean that a chimpanzee or human infant using a tool is not thinking in any way, but only that these nonlinguistic creatures are not thinking in the same way as symbolic creatures; there exist forms of cognitive activity in which they cannot engage. In my terms, perhaps what these four thinkers mean is that intersubjective and perspectival symbols that are learned and used in communicative interactions with other symbol users create the possibility of examining things from many different perspectives simultaneously, for anticipating the differing perspectives of other persons on things, and indeed for reflecting on one's own thinking from different perspectives as well. It is these kinds of thinking in which nonsymbolic creatures are unable to engage. They are unable because they do not possess the representational medium within which to conduct such dialogic and multilogic forms of mental activity; that is to say, they do not possess the representational medium that emerges as human infants begin to symbolically communicate with their fellow creatures at around 1 year of age.

Some might object to this proposal by claiming that a prelinguistic child or nonhuman primate may construe an object or situation in more than one way: one time a conspecific is a friend and the next time an enemy; one time a tree is for climbing to avoid predators and the next time it is for making a nest in. In these different interactions with the same entity, the individual is certainly deploying its attention differentially depending on its goal at that moment—taking different perspectives, if you will. But shifting attention sequentially in this manner as a function of the goal of the moment is not the same thing as knowing simultaneously a number of different ways in which something might be construed. Does a chimpanzee or a prelinguistic human infant understand that this object in front of her is simultaneously an orange, a piece of fruit, a meal, an object, a gift from a friend, a sphere, and a temptation—among other things? Does a chimpanzee or a human infant understand that this event occurring in front of her is simultaneously a fight, a chase, a social interaction, an act of retribution, an act of aggression, a tragedy, and an impending murder—among other things? If she does not, what does this imply about how she can think about,

and cognitively represent, the objects and events taking place around her?

Since Whorf 1956, the language-cognition question has focused on whether learning one language, rather than another language, affects nonlinguistic cognition. The acid test is whether learning one language rather than another language leads to some discernible difference in how individuals behave in nonverbal tasks that assess their perception of colors, or of space, or of object shapes (Gumperz and Levinson 1996). What I am talking about here—by invoking nonhuman primates, prelinguistic infants, and the like—is something different, and indeed at the moment I am agnostic about the Whorfian question in particular. What I am talking about is what Slobin (1991) calls "thinking for speaking." From this perspective, there is no privileging of nonlinguistic cognition as somehow the real thing—which we then see if language affects. From this perspective, it is preferable to simply say that cognition takes many forms depending on many factors, and one form—which is unique to the human species after 1 or 2 years of age—is linguistic cognition in which individuals structure their thinking by means of one or another historically evolved collection of intersubjective and perspectival symbols and constructions. Language does not affect cognition; it is one form that cognition can take.

To summarize, the three main claims made in this chapter—which together justify its title—are as follows:

1. The evolution of linguistic communication in the human species and the acquisition of language by human children rest crucially on uniquely human skills of social cognition. These skills enable individuals both to read the communicative intentions of others, as embodied in their symbolic behaviors, and to culturally learn those symbolic behaviors themselves.

2. When used in acts of communication, these social-cognitive skills serve to create intersubjectively understood and perspectivally based linguistic symbols, which can be used to invite other persons to construe phenomena from any one of many simultaneously available perspectives. Internalizing such acts of symbolic communication creates especially flexible and powerful forms of cognitive representation, and these then, later in ontogeny, enable metaphorical, dialogic, and reflective thinking.

3. Communicating with other persons linguistically thus leads human beings to conceptualize things and events in the world in myriad different complex ways. Without these communicative activities, human beings would have no reason to conceptualize things and events in these ways, and so they simply would not do it.

Note

Thanks to Heide Lohmann for thoughtful comments on an earlier version of this chapter.

References

Carpenter, M., Nagell, K., and Tomasello, M. (1998). Social cognition, joint attention, and communicative competence from 9 to 15 months of age. *Monographs of the Society for Research in Child Development, Volume 255.*

Gumperz, J. J., and Levinson, S. C. (Eds.). (1996). *Rethinking linguistic relativity.* Cambridge: Cambridge University Press.

Lakoff, G. (1987). *Women, fire, and dangerous things: What categories reveal about the mind.* Chicago: University of Chicago Press.

Langacker, R. (1991). *Foundations of cognitive grammar.* (Vol. 2.) Stanford, CA: Stanford University Press.

Slobin, D. (1991). Learning to think for speaking: Native language, cognition, and rhetorical style. *Pragmatics, 1,* 7–26.

Tomasello, M. (1999). *The cultural origins of human cognition.* Cambridge, MA: Harvard University Press.

Tomasello, M., and Call, J. (1997). *Primate cognition.* Oxford: Oxford University Press.

Tomasello, M., Kruger, A. C., and Ratner, H. H. (1993). Cultural learning. *Behavioral and Brain Sciences, 16,* 495–552.

Vygotsky, L. (1962). *Thought and language.* Cambridge, MA: MIT Press. (Original work published 1934.)

Vygotsky, L. (1978). *Mind in society: The development of higher psychological processes* (M. Cole, Ed.). Cambridge, MA: Harvard University Press.

Whorf, B. (1956). *Language, thought, and reality* (J. B. Carroll, Ed.). Cambridge, MA: MIT Press.