

Imitation and Human Ontogeny: Michael Tomasello and the Scene of Joint Attention

Richard van Oort

Department of English
University of Victoria
PO Box 3070 STN CSC
Victoria, BC V8W 3W1
Canada
rvanoort@uvic.ca

René Girard's mimetic theory has sometimes been criticized for being unrealistic in its implications for human ontogeny.⁽¹⁾ More precisely, it can be argued that mimetic rivalry, if it is really as fundamental as Girard maintains, would be far more evident among young children, who are universally recognized to be "imitation machines." If children learn by imitation, why are they not constantly fighting among themselves, as Girard's theory predicts? We all know the proverbial story of the children in the nursery who reach for the same toy despite the fact that there are many other toys in the room. But has anybody actually tested for mimetic rivalry among children? And what age group are we talking about? Do prelinguistic infants compete mimetically? When do children "ironize" centers of attention, like Hamlet at his uncle's court, or Tom Sawyer whitewashing his aunt's fence?

In what follows, I take a closer look at the role of imitation in human ontogeny. Much of what I say relies on Michael Tomasello's work.⁽²⁾ Tomasello is a well known comparative psychologist at the Max Planck Institute in Leipzig, and he has done a great deal of empirical research on children and nonhuman primates such as chimpanzees. My remarks will draw mainly on his account of imitation and cultural learning in his book, *The Cultural Origins of Human Cognition* (1999). In this book, Tomasello provides a fascinating picture of how children, beginning at about nine months of age, acquire higher cognitive functions, including eventually language, mathematics, and music. Tomasello's main interest, however, is not in the later period of cultural development, when children acquire advanced and culturally variable skills like long division or algebra. Rather, he is interested in what is universal in childhood development. His particular focus is therefore on the period just prior to the child's first birthday, when the child begins to engage with adults in

what he calls “scenes of joint attention.” Students of generative anthropology may be surprised to learn that Tomasello’s theory of cognitive development is very close in spirit to Gans’s reflections on the elementary linguistic forms (the ostensive and imperative).⁽³⁾ What Tomasello calls the *joint attentional scene* is the ontogenetic analogue of what Gans calls the *originary ostensive scene*.

Tomasello’s Anthropology

Tomasello begins by asking why it is that, despite our close genetic relationship to chimpanzees, we are nonetheless so different from them. Unlike chimpanzee societies, human societies are complex, and they make use of complex symbolic and material artifacts. Tomasello argues that we cannot explain this difference in purely biological terms because: “there simply has not been enough time for normal processes of biological evolution involving genetic variation and natural selection to have created, one by one, each of the cognitive skills necessary for modern humans to invent and maintain complex tool-use industries and technologies, complex forms of symbolic communication and representation, and complex social organizations and institutions” (2). Tomasello believes there is “only one possible solution” (4). This solution involves reducing the multiple differences between humans and chimps to one essential difference. Unlike chimps or other animals, humans are able to build on culture *cumulatively*.

In arguing for a fundamental constitutive difference between humans and chimpanzees, Tomasello separates himself from most primatologists and evolutionary biologists, who tend to define human culture in continuity with animal examples of what the science writer Richard Dawkins calls “the extended phenotype.” For example, birdsong can be regarded as a particularly ingenious way for transmitting bird genes. Young birds learn mating songs by imitating their parents; they do not emit them naturally or spontaneously. Likewise, chimpanzees can be said to acquire culture by imitating the tool-making and tool-using skills of their parents. Tomasello argues that such definitions of culture fail to grasp the key to specifically human culture, which is the ability to build upon culture recursively. Tomasello calls this the “ratchet effect” (37). What he means by this is that human cultural artifacts are not simply passed down unchanged from generation to generation, like the songs of a particular bird species, or the tools of the chimpanzee. They are intentionally modified by the users themselves. More specifically and more interestingly, Tomasello argues that humans participate in a collective attentional scene, the history of which stretches back to the very first scene of collective human attention. Chimps make termiting sticks as they have always made termiting sticks, but the human capacity for symbolic representation has, for better or worse, enabled the spear to evolve from pointed stick to nuclear warhead.

The skeptic may counter that this is always the case with evolution. Things tend to get more complex, as competition forces organisms and cultures to specialize into particular environmental and cultural niches. But Tomasello's point is not just that human culture is more complex, but that this complexity is built into the process of cultural transmission. Among humans, culture is modified at a rate that transcends, by several orders of magnitude, the mechanisms available to biological evolution. The rapid pace of human cultural evolution suggests that we are dealing with qualitatively different mechanisms of evolutionary change. Since humans differ very little from other primates in terms of their biology, the difference must lie in the particular mechanism used by humans to transmit culture. For Tomasello, our capacity to transmit and modify culture intersubjectively—that is, within the shared space or scene of collective attention—is what separates us from other social animals, including most notably our closest living genetic relative, the chimpanzee.

2

Though one can disagree with Tomasello on the particulars of his theory—and I will look at some of those particulars in a moment—it seems hard to deny his general point that human culture is fundamentally different from animal culture. Yet the astonishing fact is that Tomasello's position among scientists is anomalous. Why is this the case? Why do the vast majority of scientists appear so eager to deny or ignore the obvious differences between humans and nonhuman animals when it comes to symbolic phenomena such as culture and language? If I were forced to explain this curious fact, I would put it down to our natural inclination to identify with others. We can't help but explain nonhuman behavior in human terms because that is how we interpret our own behavior. Humans have an irresistible urge to anthropomorphize the world, and this urge spreads to other animals who appear to behave like us in certain respects.

I realize that this is rather counterintuitive. Surely we can expect the scientists, of all people, to be more objective. But I'm afraid I can't think of a better explanation. For a long time, we believed the whole cosmos operated in terms of human beliefs and desires. We seem to have gotten away from seeing our preoccupations mirrored in the rocks, trees, and stars. But it seems as though a residual element of this “cosmological anthropology” remains when we turn to animals, and in particular to our closest primate relative, the chimpanzee.

So what are we to do? Well, one thing we can do is listen to people like Tomasello. When it comes to comparing human societies with nonhuman primate societies, I think we have no better guide. His extensive knowledge of the literature in primatology, his own empirical research on chimpanzees and children, and above all his theoretical sophistication, make him an ideal guide for navigating the

controversial subject of human origins.

Of Chimps and Children

On the basis of his empirical experiments with children and chimpanzees, Tomasello argues that children go through a “nine-month revolution.” This is when they adopt a cognitive perspective on the world unknown to chimps or other animals as they begin identifying with the intentions and attentions of others. Before nine months, children interact with adults dyadically, for example, by imitating facial expressions, or participating in “proto-conversations” or turn-taking rituals (e.g., peek-a-boo games). Students of Girard’s mimetic theory may wish to see this pre-nine-month phase as the “Girardian” stage of human ontogeny, when infants imitate adults directly without paying attention to the world outside the model. At around nine months, however, children begin to follow the attention of adults in order to attend to external objects. This is the basis of pointing gestures. Children begin to understand that adults are beings like themselves who have goals and plans toward objects (“intentions”), and they begin to identify with those intentions by following the gaze of adults and predicting the adult’s behavior on the basis of their perception of both the adult and the adult’s relationship toward external objects. Tomasello calls this the “scene of joint attention,” and he believes it to be the fundamental basis for specifically human forms of imitation and cultural transmission. We might call this the “Gansian” phase of human ontogeny, the discovery by the child that it is a participant in the general scene of human culture.

To get a flavor of Tomasello’s argument here, consider his analysis of the widely cited example of “culture” among a group of Japanese macaques. In the 1950s an individual named Imo was observed to wash her potatoes before eating them. Gradually the habit spread, first to Imo’s closest relatives, then among the other group members. After two years, about forty percent of the troop was observed to be industriously washing potatoes. The scientists interpreted this as an example of humanlike cultural transmission, because the group’s members appeared to be *imitating* Imo’s invention of potato washing. As it turns out, however, potato washing is not quite the cultural revolution in food preparation the scientists thought they had discovered. Other individuals in other troops quite separate from Imo’s have since been observed to do the same thing. Unsurprisingly, the displeasure of chewing on sand and grit appears to be something we share in common with monkeys, who frequently can be seen to engage in the perfectly natural practice of brushing the sand off their food before chewing it. The potato washing habit is therefore better explained as an *extension* of this natural brushing behavior. Given exposure to water and sandy potatoes, sooner or later the monkey will discover that washing the potato with water is a more effective sand-removing technique than simply brushing it. But the real clincher for Tomasello is the fact that

the rate at which the habit of potato washing spread within Imo's troop remained constant throughout the two-year observation period. If individuals really were imitating the behavior of their fellow macaques rather than relying on individual trial and error, one would expect the rate of transmission to increase dramatically as the number of potato washers increased. But this was not the case. Tomasello's conclusion is that the macaques were not so much imitating Imo's behavior as being led by her to favorable circumstances in which each individual could *discover for itself* the elegant beauty of potato washing. The fact that Imo's closest relatives adopted the habit first is consistent with this hypothesis. As they foraged with Imo, they were the most likely members to be in the same vicinity of water and sandy potatoes. They therefore were also the first to discover, one by one and by individual trial and error, the handy trick of potato washing.

Tomasello calls this type of social learning *emulative* rather than imitative. The difference is that in emulative learning the disciple focuses not on the model's particular behavior but on the objects with which the model is interacting. Chimpanzees, for instance, are very good at observing other chimpanzees interact with objects, but they do not then imitate the other's behavior with respect to the objects involved. Rather, they are led to interact with the objects and discover for themselves the *natural affordances* of the particular objects attended to. For example, a young chimp may observe its mother crack a nut with a stone. It will then pick up a stone and discover that the stone makes a pretty good hammer. What the chimp has learned is not a particular behavior (the mother's technique of nut cracking), but a fact about stones, and more precisely, a fact about the impact of stones on nuts. This is something that is learned by the chimp in its interaction with the stone and the nut, not by imitating its mother's gesture toward the nut or stone. The chimp does not oscillate its attention between mother and the objects involved, in a conscious effort to reproduce her particular gesture. It therefore has not learned a new behavior by imitation. Hammering is something the chimp can do individually, given the natural affordances of the objects involved. Provided with a stone and a nut, and given the general primate capacity for grasping objects (Thank God for the opposable thumb!), the chimp discovers how to crack a nut. The behavior is emulative rather than imitative because, as Tomasello puts it, the chimp "focuses on the environmental events involved—the changes of state in the environment that the other produced—not on a conspecific's behavior or behavioral strategy" (29).

In order to test his theory that chimpanzee learning is emulative rather than imitative, Tomasello devised a series of ingenious experiments testing both two-year-old children and chimpanzees. The experiments involved getting the subjects

to imitate the behavior of a model. Tomasello describes how the children almost always insisted on imitating the behavior of the model, no matter how bizarre or inefficient it was. For example, if the model switched on a light by using her head, so would the children. Or if the model used a tool in an extremely inefficient fashion to reach an object, the children would use the tool in the same inefficient manner, despite the fact that the natural affordances of the objects presented a much easier way to do the same task. The chimps, on the other hand, simply experimented with the objects no matter which way had been demonstrated to them beforehand.

Evidently, the children were imitating the model whereas the chimps were attempting to emulate the outcome of the experiment independently of the model's particular behavior. Whereas the children were focused on the model's *behavior toward the goal*, the chimps were focused on the outcome of the experiment. The difference is important because it explains why chimpanzees have such great difficulty learning to use symbols. The ability to separate behavior from outcome is necessary before the model's gesture toward the object can be transformed into a genuine symbol that designates or "means" the object. What Tomasello's experiments strongly suggest is that children, but not chimps, are predisposed to focus on the model's *behavioral stance toward the object*. They are entering into the model's particular intentional stance toward the object.

Joint Attention

Key to Tomasello's ideas about language acquisition among children is his hypothesis concerning the joint attentional scene. Before nine months, children interact with the world much as other primates interact with the world. That is, they are aware of the objects around them and of other individuals interacting with those objects, but they never enter into the other's intentionality toward those objects. If a pre-nine-month-old child is playing with an object and an adult walks into the room and says, "Look, let's play with this!" while holding out a toy car, the child may reach out and grab the car and start sucking on it, or manipulating it, or whatever, but it pays no attention to the adult's intentions toward the toy. It pays attention either to the toy or to the adult, but not to the relationship between adult and toy. In other words, the child does not think to itself, "Oh, mommy wants to show me this new toy," or even, "This person wants to show me this thing." On the contrary, if Tomasello is right, it does not even "think" at all, at least not in the way an adult thinks, which according to Tomasello is an internalized version of the kind of joint attentional scenes children first experience at nine months. Thus, in grabbing the toy the pre-nine-month-old child does not look from mommy to toy and back to mommy again. The toy is simply another object to interact with, but it receives no further significance beyond the child's own interest in it. Tomasello's argument is that chimps never really go beyond this "egocentric" stage of understanding objects, which is why primatologists never observe chimps pointing

in the wild. That is, they do not distinguish between my intention toward objects and your intention toward objects, so there is no point in trying to get you to pay attention to my attention toward the object. As far as chimps and pre-nine-month-old children are concerned, there is only the point of view of the self, into which all other perspectives are innocently absorbed.

At around nine months of age, however, children begin to engage in what Tomasello calls *joint attentional scenes*. Initially, this begins with simple checking on the attention of an adult in relation to an outside object, but it quickly evolves into gaze following, when the child looks from adult to where the adult is looking, and to acts of pointing, when the child tries to direct the attention of an adult to some external object. What Tomasello is keen to stress in this ontogenetic “revolution” is the fact that the scene is fundamentally triadic in structure: “Joint attentional scenes are social interactions in which the child and the adult are jointly attending to that third thing, for some reasonably extended length of time” (97). The child’s attention shifts between the adult and the object to which both adult and child are attending. In this collective sharing of attention toward a central object, Tomasello sees the roots of symbolic culture, including language, symbolic play, and ritual.

Joint attentional scenes, however, are not examples of language, at least not language in the sense usually intended by philosophers or linguists.[\(4\)](#) They are rather the minimal condition of language. On the other hand, nor are they simply perceptual events of the kind that nonhuman primates and other animals engage in. All animals, including of course all humans, perceive the world around them and are able, on the basis of those perceptions, to form sensory-motor representations that allow them to anticipate events in the world, including the actions of other conspecifics. However, these anticipations are still perceptually based, in the sense that they are individually learned image schemas or sensory-motor representations. To borrow a usage from the evolutionary anthropologist Terrence Deacon, these perceptual and sensory-motor representations are *indexical*. They are based on the capacity of all animals to form categories of perceptual events, including categories of communicative events, such as the widely publicized example of vervet monkey distress calls.[\(5\)](#)

The scene of joint attention is quite different. Indeed, Tomasello claims that it bridges the gap between perceptual representation (which we share with all animals) and language (which only we possess). What differentiates the joint attentional scene from language in the narrow sense employed by linguists is the fact that language “abstracts” from the scene to include only its most portable aspect, which is the symbol or word itself. On the other hand, what differentiates the joint attentional scene from perceptual events is that it includes “only a subset of the child’s perceptual world” (97). That is, the joint attentional scene focuses the

child's attention on a central object, against which all other perceptual objects and events become background or "periphery."

Let me emphasize this difference between perceptual events and the joint attentional scene. Animals can of course focus their attention on discrete objects or events in the world, as when a cat tracks the movements of a mouse, or a chimp warily eyes the presence of a male rival. In the joint attentional scene, however, the child is not merely paying attention to the object, but to *someone else's attention toward the same object*. In other words, the child grasps that the significance of the object is mediated by the attention of the other. It is this capacity to separate the other's intention-his or her internally represented goal-from the perceptual reality of the object that distinguishes the joint attentional scene from otherwise superficially similar perceptual scenes. The child learns that the adult's intention to the object is distinct from its own intention toward the same object. Moreover, in making this distinction between self and other, it lays the foundation for participating in an intentional relation that is truly collective or intersubjective, because in recognizing the difference between the other's intentionality toward the object and the object itself, the child learns to take a perspective on the object distinct from its own. The child is now imitating a particular intentionality toward the object that is transposable to other scenes in which the object may appear.

Tomasello calls this "role reversal imitation" (105) because it implies that the child is able to grasp that an adult's intentional stance toward an object is something that can be adopted by the child itself. This is something we do all the time-indeed, whenever we use language. For example, suppose you tell me that the peculiar thing on your dining room table is a "grazza." Later, when my wife walks in the room and makes a face while staring at the peculiar object on the table, I turn to her and say, "Oh, that's a grazza." I have not merely imitated the word, I have also reproduced your intentional stance toward it. That is, I have recreated the joint attentional scene by adopting your perspective, and this time I have reversed the roles because I am now instructing someone else, as you had instructed me before.

4

A skeptic might object that these joint scenes of attention are not so very different from the attentional scenes other animals engage in. Animals are not rigidly tied to the same perceptual construal of a particular object. A tree may represent a number of different things. Depending on the context, it may represent an escape route, a nesting site, or the location of food. But Tomasello's point is not that perceptual scenes are inflexible but that they are always tied to the natural affordances of the objects and that these affordances are discoverable by the individual's dyadic interaction with the object. At no point does the chimp seek confirmation *from another conspecific* in order to see that the tree is a nesting site

or an escape route or source of food. It can discover these things for itself. Furthermore, it is impossible for the tree to represent all these things *at the same time*. The chimp does not choose between different representations of the tree that it can hold in its mind simultaneously. Rather, the representation of the tree remains a function of the chimp's particular goal, which is either to eat, escape, or sleep. As Tomasello says, "the animal is attending to different affordances of the environment depending on its goal" (126).

In the case of symbolic attention, however, the goal is not defined by the practical affordances of the environment, but by the attentions of both individuals in the attentional scene. When a child points to a tree, the goal is to secure the adult's attention to the same object. And this is, in the end, what language does. It secures the other's attention toward some external object or, in the case of declarative sentences, some external idea or "signified." When it comes to specifically human cognitive functions, what is primary, as Emile Durkheim saw, is the *social relation*. It is the latter that mediates our more basic indexical perceptual and sensory-motor functions. The latter are basic functions that we share with all other animals. It is the mediation of these functions by the joint attentional scene that distinguishes human from animal cognition.

In a manner reminiscent of Deacon's theory of symbolic reference, Tomasello argues that language emerges as a negation—or, in Hegelian terms, a transcendence—of more basic perceptual and sensory-motor representations. In order to construe an object in symbolic terms one must impose an intersubjective relation onto a perceptual relation. That is, one must enter into a joint attentional scene that can define the object "arbitrarily" in terms of each participant's shared attention to the object. This creates an intersubjectively shared "space"—a period of deferral, if you like—between the two participants in which the object is "centralized" as the shared focus of attention. Rather than seeing the object as a function of my individual biological needs (e.g., as a place of rest, escape, or food), I see it as a function of your attention, which—as students of mimetic theory well know—may be in conflict with my own pragmatic designs on the object. Gans, following Durkheim and Girard, suggests that the originary act of symbolic designation is an act of *sacralization*. The designation of the object as sacred is not something that can be understood on the basis of the natural affordances of the object. On the contrary, it is an arbitrary imposition, in the sense that its "functionality" is given by the intersubjective relation itself. The object is now attended to as a function of the symbol used to designate it. As Tomasello points out, this is the basis for the perspectival nature of symbols. Language is used in order to get someone to attend to the world in a certain way. In designating an object symbolically, I first have to decide what symbol to use. For example, I could call a tree, *that tree, the oak, the tree in my backyard, the monstrosity that blocks my sunlight*, or any number of

things. But how I choose to construe it is, in the end, always a function of how I think I can best get you to attend to it with me. That is, I choose between different symbols by simultaneously monitoring your attention to the object. This is in fact how children acquire language and this is also what we mean by imitation in the specifically human context. Symbols are “attention getters.” They are the tried and tested means passed down from previous generations of language users for participating in joint attentional scenes. “In imitatively learning a linguistic symbol from other persons,” Tomasello says, “I internalize not only their communicative intention (their intention to get me to share their attention) but also the specific perspective they have taken. As I use this symbol with other persons, I monitor their attentional deployment as a function of the symbols I produce as well, and so I have at my disposal both (a) the two real foci of self and communicative partner and (b) the other possible foci symbolized in other linguistic symbols that might potentially be used in this situation” (128). As this passage implies, Tomasello’s theory is aimed at explaining the ontogenetic pathways of the child’s entrance into the “mimetic triangle” of human culture.

Objects as Symbols

One thing I would like to emphasize in Tomasello’s account of human ontogeny is his view that the child’s symbolic interpretation of intentional objects follows, rather than precedes, the child’s acquisition of ostensive words in scenes of joint attention. At first, this might seem rather counterintuitive. We tend to think of words as horribly abstract whereas objects, even symbolic objects, are tangible and concrete. At least you can manipulate a toy car, even if it is “only” a representation, a model of the real thing. The word *car*, on the other hand, is by comparison a very abstract thing, little more than a puff of air or, in the case of writing, black dots on the page or computer screen.

But that is precisely the point. In order for the child to move from concrete perceptual and sensory-motor representations to abstract symbols, it needs to override all those perceptual and sensory-motor associations it has learned in the first nine months of its life. But this is a very hard thing to do if your raw material consists of graspable objects with all kinds of preexisting intentional and natural affordances. An interesting discovery of Tomasello’s experiments was that it is very hard for children under two to intentionally interpret a cup as a hat, or a pencil as a hammer, for example, by putting the cup on their heads, or hammering with the pencil. These symbolic interpretations are difficult for the child because the objects already possess clear cut intentional affordances. The cup is for drinking, the pencil for drawing. These more basic perceptual and sensory-motor representations tend to override the child’s relatively undeveloped capacity for symbolic association and metaphoric thinking.

In a particularly poignant experiment, Tomasello demonstrates how hard it is for children under two to interpret nonarbitrary objects in purely symbolic terms. Children aged eighteen to thirty-five months were asked to give the experimenter an object. In the first stage of the experiment, the experimenter simply asked for the object by name. All the children responded appropriately. In the second phase, the experimenter asked for the object by holding up a toy replica of the object (e.g., holding up a toy hammer in order to get the real hammer). Interestingly, the children under twenty-six months had extreme difficulty with this task. They reacted instead by reaching for the toy held up by the experimenter. Children over twenty-six months, however, had no difficulty interpreting the toy object symbolically as a request for the object represented. Tomasello suggests that the reason the task is so difficult is “that the younger children engaged with the toy object as a sensory-motor object,” and that this engagement prevented them from interpreting the object as a symbol of *something else*, namely, the real object the experimenter was requesting (86). This is an interesting finding because it suggests that symbolic iconicity, far from being a natural stepping stone toward language, is in fact something children grasp only once they have already mastered ostensive words (e.g., “Juice!” “Dog!” “Tree!” etc.). I interpret this as additional evidence that, phylogenetically speaking, we have no choice but to interpret the origin of language and culture as a radical break from preexisting animal forms of culture and communication. There is no shortcut from perceptually based modes of iconicity and indexicality to genuine symbols. The unbounded human capacity for metaphor and other forms of symbolic analogy begins with the “vertical” separation between the central object and the intersubjectively shared sign. The aborted gesture of appropriation that defers the indexical relation between subject and object is the “humble” beginning of humanity—the “little bang,” as Gans puts it.⁽⁶⁾ The key ingredient of the originary scene is the minimal symbolic sign.

Human Phylogeny and the Joint Attentional Scene

Before I conclude, I would like to make a small criticism of Tomasello’s account of human phylogeny. In general, I agree with most of what he says on this topic. I agree that evolutionary biologists, anthropologists, and psychologists have, by and large, neglected the “cultural” factor in human evolution. Instead, there has been too much emphasis on the genetic or biological side of things. Tomasello rightly disputes the simplification of this issue into an inflexible dichotomy between biology and culture. As he points out, the dichotomy doesn’t really exist. What “exists” are rather different ideas about history. There is phylogenetic time, which is the perspective of biological or genetic evolution. There is historical time, which is the perspective of human beings reflecting on their relationship to past culture and in

particular to past ideas that other human beings have had about themselves and the world around them. And, finally, there is ontogenetic time, where biology and history each play an indispensable role in the formation of the human individual. Tomasello thinks that evolutionists have favored looking at human evolution in phylogenetic or biological terms because it just seems easier and more elegant. It is much simpler to posit a genetic event as the cause of something because genetic events are more manageable than cultural events, which tend to be rather messy and imprecise affairs. Hence the temptation to see human cognition in terms of a number of discrete "modules" that are genetically wired to produce different "types" or "categories" of cognition. So, following this line of argument, there must be separate cognitive modules for perceiving objects, for knowing persons, for recognizing number, for acquiring language, and so on and so on. Obviously, this sort of thinking is not very rigorous and it is easy to see how it can quickly get out of hand. Do we need a module for chess? What about the God gene?

Tomasello rightly cuts through the confusion implicit in this kind of thinking. His expertise on human ontogeny allows him to see that the cultural side of the story is really much more important than the typical evolutionist admits. This is what is so valuable about Tomasello's account. He sees the importance of culture—the mediation of the child's attention by the joint attentional scene—in changing the biological pathways of the child's cognitive development, as it goes from perceiving the world much like other primates do, to seeing it in terms that only humans do. However, I have to point out that Tomasello doesn't always follow his own advice. In his concluding section, after he criticizes the modular theorists for their ad hoc practice of explaining human cognition in terms of multiple discrete genetic events, he goes on to propose a genetic event of his own:

My attempt is to find a single biological adaptation with leverage, and thus I have alighted upon the hypothesis that human beings evolved a new way of identifying with and understanding conspecifics as intentional beings. We do not know the ecological pressures that might have favored such an adaptation, and we can hypothesize any number of adaptive advantages it might have conferred. My own view is that any one of many adaptive scenarios might have led to the same evolutionary outcome for human social cognition, because if an individual understands conspecifics as intentional beings for whatever reason—whether for purposes of cooperation or competition or social learning or whatever—this understanding will not then evaporate when that individual interacts with conspecifics in other circumstances. (204-5) It's too bad Tomasello doesn't follow his own advice. Here he gets it exactly backwards. Understanding other conspecifics as intentional beings like oneself is not a "genetic event" because there is no conceivable reason why, genetically speaking, such an event should occur, as he himself admits. Or, to put the same point differently, there are so many reasons

why the capacity to identify with others is beneficial *from the perspective of those who already have it* that the number of plausible originary scenarios is limitless. But the real issue is not the infinite number of scenarios we can imagine for causing individuals to identify with one another. It is the presence of the scene of identification itself, which takes place, as Tomasello elegantly shows, in the joint attentional scene between parent and child. It is the joint attentional scene that must be explained, not the “genetic” ability to identify with others. The “selection pressure” to see other beings like oneself is given by the structure of the (joint attentional) scene itself. The mimetic “pressure” to maintain attention on a central object *that is also the center of attention of the other* leads the subject to begin to identify with the other’s internal representation of the object, and vice versa. Tomasello’s detailed observations of young children provide empirical confirmation of this hypothesis. Where Tomasello trips up is when he translates his theory of human ontogeny into a hypothesis for a single genetic event in human phylogeny. But the joint attentional scene is irreducible to a genetic event. The origin of humanity is the first scene of joint attention, the minimal mimetic triangle of the originary hypothesis.

Conclusion

I began this essay by asking why it is that Girard’s mimetic theory has so little to say about human ontogeny despite the fact that children are “imitation machines.” I now wish to propose an answer. The reason is that Girard’s theory does not distinguish systematically between two kinds of imitation, imitation of a model and imitation of an object. For Girard, imitation is always ultimately imitation of someone else. This is why he associates the mimetic crisis with the loss of difference between subject and object. In a mimetic crisis, the rivals becomes so obsessed with each other that they no longer grasp that what they are designating is something external to both of them, the central object of the joint attentional scene. Girard’s idea of imitation is almost wholly based on this idea of dyadic imitation.

What Tomasello’s account suggests, however, is that dyadic imitation is a necessary ontogenetic step toward gaining access to the intentions of the other. This is a far more constructive notion of imitation. In paying attention to the behavior of the other, I am not merely imitating the other, but entering into the other’s perspective on the world. As Tomasello suggests, I am identifying with the other. But this identification would not be possible without the presence of a third element, which is the object to which we are both paying attention. In *Signs of Paradox*, Gans emphasizes the necessity of including the object as an indispensable third element in the mimetic relation, despite the fact that this seems to contradict Girard’s original idea of the mimetic crisis:

Why should the intensification of mimesis lead the subject away from the other's behavior toward the object to which it is directed? This movement reflects an internalization of the model's motivation, the self's closer assimilation to the other's own reality. The more closely I imitate my model's goal-directed action, the more I share the goal of this action, which is not located in the action itself but precisely in the external object. [. . .] Whence the apparent paradox that as imitation becomes more intense, it prefigures the triangular structure of human representation, focusing less on the model's behavior and more on the object to which it is directed.⁽⁷⁾ Tomasello's account of human ontogeny confirms Gans's theory of imitation. The child learns to imitate not merely the external action of the other, but the other's internal goal as well. The joint attentional scene is the basis for child's acquisition of wholly abstract objects, such as the ideas or "signifieds" of declarative language.

In proposing that mimetic rivalry dissolves the mimetic triangle between self, other, and object, Girard reverses the passage from nature to culture, or, in Tomasello's ontogenetic terms, from the infant who interprets the world much like other primates do, to the child who interprets the world symbolically in terms of the joint attentional scene. But this reversal is not really a reversal of the originary passage from nature to culture. It is a historical renewal of it. In the case of children, this historical renewal is experienced ontologically, in the sense that their identity as individuals is dependent upon their cognitive acquisition of culture, as they begin to participate in scenes of joint attention and internalize the mimetic configuration between self, other, and center of attention.

One can agree with Girard that too close a focus of attention on the other can lead to forms of mimetic rivalry that may become counterproductive. But Tomasello's research suggests that imitation is in fact a far more flexible phenomenon than Girard acknowledges. The child's identification with someone else is possible only once the child can engage in scenes of joint attention. But entry into the latter is also the source of an immense cultural productivity that can turn mimetic rivalry into any number of more "peaceful" solutions. Indeed, the conflict between self and other—for example, in the scenes of sibling rivalry examined by Girard in his analyses of literary and religious texts—frequently provides the motivation for constantly renewing the joint attentional scene, for instance, in an extended dialogue in which the interlocutors attempt to see the other's point of view and vice versa. Tomasello cites research that young children who have siblings are more likely to identify with other points of view, because they have learned from an early age to engage in joint attentional scenes in which their desires have been in conflict with someone else's. The joint attentional scene provides the opening to ultimately

limitless forms of negotiation and deferral as participants seek to engage each other in their shared and-thanks to their status as co-equals in the scene-sharable perspectives on the world. The generativity of this scene is the core of Tomasello's theory. His reflections on the ontogeny of this scene, as children are encouraged to participate in it by their parents, is a powerful example of originary thinking in the social sciences.

7

Notes

1. I'm afraid I don't have a specific reference for this criticism of Girard. Nor do I know if Girard himself has made any reference to imitation in human ontogeny. Nonetheless, in casual conversation I have repeatedly heard the claim that children demonstrate Girard's theory rather well. Recently, Matthew Taylor raised the issue on the GA blog. He pointed out that at least one social scientist has disputed the Girardian claim about mimetic rivalry among children. See <http://dev.cdh.ucla.edu/GABlog/?p=15#comments>. I would like to thank Matt for raising the question. I hope this article goes some way to providing an answer. [\(back\)](#)
2. See, in particular, Michael Tomasello, *The Cultural Origins of Human Cognition* (Cambridge, MA: Harvard University Press, 1999) (hereafter cited in text). [\(back\)](#)
3. See, for example, Eric Gans, *Originary Thinking: Elements of Generative Anthropology* (Stanford: Stanford University Press, 1993), especially chapter 4. [\(back\)](#)
4. This seems to be the source of Derek Bickerton's objection that Tomasello's theory would be better served by postulating that language rather than intentionality were the originary basis of human culture. I tend to agree with Bickerton that Tomasello risks reifying the notion of intentionality, which he seems to regard as a purely biological phenomenon. It is more minimal to assume that once the joint attentional scene has emerged-protolanguage, in Bickerton's sense-then we can assume that specifically human forms of symbolic intentionality emerge with it. I think, however, that Tomasello's idea of the joint attentional scene already implies that human intentionality is a scenic phenomenon. I therefore will not dwell on his less parsimonious claim for the causative role of biological intentionality. For Bickerton's objection, see his response to the article by Michael Tomasello et al., "Understanding and Sharing Intentions: The Origins of Cultural Cognition," *Behavioral and Brain Sciences* 28 (2005): 675-735. [\(back\)](#)

5. See Terrence Deacon, *The Symbolic Species: The Co-Evolution of Language and the Brain* (New York: Norton, 1997). See also my “Cognitive Science and the Problem of Representation,” *Poetics Today* 24 (2003): 237-95. [\(back\)](#)
6. See, for example, Eric Gans, “The Little Bang: The Early Origin of Language,” *Anthropoetics* 5.1 (1999): 6 pp. [\(back\)](#)
7. Eric Gans, *Signs of Paradox: Irony, Resentment, and Other Mimetic Structures* (Stanford: Stanford University Press, 1997), 23. [\(back\)](#)