

Along these lines, Povinelli (this volume) argues convincingly that chimpanzees and other higher primates are skilled at detecting structure in action but lack the inferential ability necessary to interpret this structure in genuinely intentional terms. If so, then divining the intentional significance of one another's actions may be a uniquely human preoccupation. We join with Povinelli in hoping to focus future inquiry on the precise nature of the relation—in both evolution and development—between skills for low-level behavior analysis and inferences about intentions.

Such questions aside, our findings clarify that infants as well as adults readily detect structure in action that is relevant for drawing inferences about intentions. In our view, low-level action analysis skills likely enable such structure detection. These basic parsing abilities gain infants access to appropriate units within complex, continuous behavior, and help adults similarly process novel action sequences for which they lack a ready motivational explanation. In this sense, structure-detection skills play a crucial role in potentiating our lifelong pursuit of interpersonal understanding.

#### Acknowledgments

This work was supported by a National Science Foundation Graduate Research Fellowship to the first author and by a NSF New Young Investigator award (No. 9458339) and a John Merck Scholars Award to the second author. Our thanks to Janet Astington, Rebecca Brand, Diego Fernandez-Duque, Bertram Malle, Lou Moses, Daniel Povinelli, and Mark Sabbagh for helpful comments.

#### Note

1. The same pattern of results was found for both long and short action sequences. For clarity, only the findings from the short action sequences are presented here.

Intention → Intentionality  
Malle, Moses + Baldwin  
MIT Press, 2001

## 10

### Desire, Intention, and the Simulation Theory

Alvin I. Goldman

A common goal of philosophy of mind and psychology is to understand how people acquire, represent, and deploy such mentalistic concepts as desire, intention, and belief. Developmental psychologists call this subject *theory of mind*. Although this label may not commit them to the so-called *theory-theory*, almost all developmentalists (with the major exception of Paul Harris) in fact favor the theory-theory approach to the subject. A greater division of opinion prevails among philosophers, with a dedicated coterie advocating a competing approach known as the *simulation theory* (Gordon 1986; Heal 1986; Goldman 1989; Currie 1995). In this chapter I argue that the vaunted theory-theory (TT) is not on very secure ground, that it suffers from serious weaknesses not adequately appreciated by many of its advocates, and that the simulation theory (ST)—at least, its “introspectionist” variant—promises to be a cure for many of these ills. I discuss these issues with special attention to desires and intentions.

#### Theory-Theory and Conceptual Change

The standard version of TT is a combination of two theses: (1) that common-sense psychological concepts are theoretical concepts, similar in all relevant respects to the theoretical concepts of physical science, and (2) that people detect psychological states in themselves and in others by making theoretical inferences. In trying to defend these two theses, developmentalists typically appeal to multiple stages in children's grasp of mentalistic concepts. First, they find evidence of performance changes in mental-state attributions. Second, they interpret these performance changes as changes in children's conceptual repertoire concerning mentalistic states. Third, they

infer from these supposed changes in conceptual repertoire that the concepts in question must be theoretical ones.

The orthodox position is that children move from a nonrepresentational understanding of mind to a representational understanding. This orthodoxy, however, has two variants. On one view, children's grasp of the very same mentalistic term changes over time. For example, they initially understand *belief* in a nonrepresentational fashion, but later (e.g., by age 4) they understand *belief* representationally (Gopnik 1993). On the other view, there may be no single mentalistic term that is conceptualized differently at different times. Rather, before a certain early age only nonrepresentational concepts are acquired, and representational concepts are mastered later. Bartsch and Wellman (1995a) deny that belief is first grasped nonrepresentationally and later grasped representationally. Nonetheless, they retain allegiance to the basic progression from nonrepresentational to representational concepts in claiming that only nonrepresentational concepts (such as the "simple," nonrepresentational concept of desire) are grasped before age 3, and the representational concept of belief is acquired thereafter. Similarly, Perner (1991) appears to hold that no truly representational concept is acquired before age 4, and up to that age children do not really understand any concept of belief. I shall examine these two variants of the orthodox TT position separately, identifying difficulties for each.

The first view is supposedly supported by a change in children's competence at false-belief tasks. It is assumed that this change can be attributed to a change in the *conceptual grasp* of belief, from a nonrepresentational to a representational understanding. The tenability of this assumption, however, has been undercut by accumulating evidence that young children's poor performance on false-belief tasks is not a *conceptual* deficit. A number of studies suggest that 3-year-olds' fragile memories for narratives are central to their poor performance on the traditional false-belief tasks. Three-year-olds do succeed on these tasks when they receive an aid to memory, or when they consolidate the required information in memory by hearing a story line twice (Mitchell and Lacohee 1991; Lewis, Freeman, Hagestadt, and Douglas 1994; Freeman and Lacohee 1995). Clements and Perner (1994) find evidence of an implicit grasp of false belief even when explicit answers to false-belief questions do not reveal understanding.

Three-year-olds also demonstrate an ability to recognize false beliefs associated with unfulfilled intentions (Moses 1993).

Another alternative to the conceptual deficit explanation of 3-year-olds' poor performance on false-belief tasks is the executive-control explanation offered by Russell. In the "windows task," Russell, Mauthner, Sharpe, and Tidswell (1991) found that in 3-year-olds and in autistic children salient knowledge of physical reality overwrote knowledge of mental reality. These children could not help pointing to a box that visibly contained a chocolate rather than to an empty box, even though they knew that it was in their interest to point to the empty box. Their failure to point to the empty box perseverated over 20 trials. This failure of 3-year-olds and autistics, in contrast with normal 4-year-olds, was associated with comparable performances on false-belief tasks, suggesting that the latter task is also difficult because it requires subjects to inhibit reference to a salient object.

Finally, Bartsch and Wellman's (1995a) extensive data provide no evidence that 3-year-olds have only a nonrepresentational construal of belief. They write (*ibid.*, p. 57):

... our data provide no evidence that a representational understanding of beliefs is a significantly later achievement, following only on the heels of an earlier "connections" misconstrual of beliefs. Instead, the data suggest that very soon after children begin to talk about thoughts at all, they discuss a variety of distinctions among mental contents and states of the world, including some that seem to presuppose a representational understanding of mental contents as separate from but about the world.

Let us turn, then, to the second variant of the orthodox TT position, which holds that during an early period only nonrepresentational mentalistic concepts are grasped. This view is defended in detail by Wellman and collaborators (Wellman 1990; Wellman and Woolley 1990; Bartsch and Wellman 1995a), who argue that the concept of desire is acquired much earlier than that of belief—roughly at 2 years—but is a nonrepresentational concept. Like Flavell (1988) and Perner (1991), Wellman and collaborators say that 2-year-olds have only a simple "connectionist" grasp of wanting, involving a relation between the target person's mind and some object or state of affairs *in the world*.<sup>1</sup> Wanting is not construed as involving a mental representation of objects or states of affairs. As Wellman puts it (1990, p. 211), "the conceiver can simply think of the target person as having an internal longing for an external object."

Let us examine more closely this story about desire, which I shall call *the world-related story*.<sup>2</sup> Two-year-olds, say Bartsch and Wellman, do not attribute any mental content to the target agent to whom the desire is imputed. "Whatever contents are involved are out there, in the world, rather than in the person's mind." (Bartsch and Wellman 1995a, p. 151) Can this account of a simple form of intentionality work? Bartsch and Wellman acknowledge that, even for 2-year-olds, desires are not restricted to physical items such as apples or toys. Objects of desire include actions and states of affairs (even states of affairs that do not currently obtain). A 2-year-old might construe someone as wanting a future state of affairs, such as a full cup of milk being before her.

To assess the prospects of the world-related story, it behooves us to begin with the type/token distinction. As applied to actions, it should be drawn as follows. Actions can be considered either as types or as tokens. *Winning a race* is a generic *type* of action, which different people can instantiate (realize, token) on different occasions. The general type, winning a race, is something like a universal (in the philosopher's sense), and therefore it is not something in the world. But different *tokens* or realizations of this type *are* in the world: Sergio's winning the race in Italy last Sunday, Kristin's winning the Boston marathon last summer, a certain thoroughbred's winning the fourth race at Hialeah. Of course, action tokens are in the world only if they are actual, not merely possible or potential (though their actuality might reside in the future). The same points apply to states of affairs.

Now, under the world-related story about desire, what token is the object or "relatum" of an action desire when that desire goes unsatisfied or unfulfilled? Suppose Sammy wants to win the 50-yard dash at school today, but actually he loses. What action token *in the world* is Sammy's desire related to, according to the world-related story? Since there is no actual token of Sammy's winning that race, it is far from obvious what a plausible candidate might be. It wouldn't be Sammy's winning the 50-yard dash last week, since Sammy's desire of today is not directed at last week's victory. Nor would it be somebody else's winning a 50-yard dash today; Sammy did not desire anybody else's victory, only his own. There seems to be no good candidate for an action token in the world that is the relatum of Sammy's desire. This is precisely why philosophers of intentionality have argued that

intentional mental states, like desires, wishes, or intentions, are relations to some sort of mental contents or representations.

Bartsch and Wellman (1995a, pp. 152–153) try to address this issue, using an example of a target person wanting a cup of milk when there is no visible cup of milk in the wantor's neighborhood:

... since the milk is not visible, where does the child's construal of the target person's desire for it come from? On this proposal, it comes from the 2-year-old's knowledge of the world. The child simply knows . . . that there are such things in the world (even if none are currently visible), and he attributes to the target person a simple desire for those real-world contents. The 2-year-old's knowledge of the world fills in certain gaps, but what the 2-year-old attributes to people is a desire for real objects, actions, and events in the world.

This does not resolve the problem. Suppose the 2-year-old attributes to the target a desire to drink a cup of milk. It does not help to be told that the attributer knows that there have been various milk drinkings in the past. The question remains: Which actual milk drinking in the world does the attributer take the target's desire to be related to? (Assume, as before, that the desire does not get fulfilled.) No answer seems to be forthcoming.<sup>3</sup>

The drift of my argument is that no good sense can be made of desire without positing something like mental representations as the relata of desire.<sup>4</sup> Of course, this does not mean that mental representations are *what* people want in the sense that the existence of such a representation automatically fulfills or realizes their desire.<sup>5</sup> Rather, mental representations express or constitute the *contents* of their desires—contents that may or may not be realized. Real-world entities cannot cover the relata of desire in all cases, including cases understood by 2-year-old attributers.

It may be responded that, even if no coherent, comprehensive, world-related story can be told about desire, that does not prove that 2-year-olds do not have such a conception. Maybe they are confused; they just haven't thought the matter through very clearly. I agree that this is possible. On the other hand, what evidence convicts 2-year-olds of such confusion? *We could* ascribe a confused model of desire to them; but what compels us to withhold from them a representational model, and what forces us to ascribe to them a confused, world-related model?<sup>6</sup>

Indeed, there seems to be every reason to ascribe a representational model to 2-year-olds. It is universally agreed that an understanding of false belief is a sufficient indicator of a representational concept of belief. Unfulfilled

desire (or intention) is precisely analogous to false belief; the only difference is "direction of fit" between mind and world (Searle 1983). Since Bartsch and Wellman grant that 2-year-olds understand that desires can go unfulfilled, the parallel conclusion to draw is that 2-year-olds have a representational concept of desire. True, desires do not represent in exactly the same way that beliefs do. Desires do not depict the way the world is; they only represent how the world *should be* (as judged by the agent's preference). But this simply is the direction-of-fit difference between desire and belief. It is not a reason for a theorist to withhold representational status from desire, unless the theorist confuses the generic idea of representation with a particular subspecies of it (*viz.*, *depictive representation*).<sup>7</sup>

An additional problem with the second version of the orthodox TT approach arises from the doctrine of *meaning holism* that is usually associated with TT. TT standardly includes the idea that mentalistic terms, being theoretical terms, are defined or understood in terms of the theory in which they are embedded. This view is inherited from traditional philosophy of science. When the theory changes, the meanings of the terms change—at least when there are changes in the laws or explanations in which those very terms appear. Even with the mere expansion of a theory (as opposed to its revision), the terms of the expanded theory will express different concepts than those of the predecessor theory. Let us apply this to the story about desires and beliefs that Bartsch and Wellman (among others) tell. When the child has a concept of desire without any accompanying concept of belief (between ages 2 and 3, according to Bartsch and Wellman), the concept of desire will be defined in one way. When the child then acquires a concept of belief (at age 3, roughly), the concept of desire will have to undergo some sort of change, because it will no longer be embedded in the same theory. The theory will at least be expanded, if it is not revised. So the 3-year-old cannot have the same concept of desire as the 2-year-old, according to TT. What is the evidence, however, that such a change occurs? I do not know of any such evidence. This raises doubts about the conceptual doctrine of TT, in comparison with alternative doctrines which would not require old concepts to change merely because new ones are added. Such an alternative would go naturally with introspectionist versions of ST unwedded to meaning holism.<sup>8</sup>

It is instructive to appreciate the counterintuitiveness of meaning holism when applied to conceptual expansions. Suppose a child already has the

concept of paper. She has drawn on paper with pencils and crayons, she has glued things to paper, and so forth. One day she is introduced to pen and ink, and shown how to use them on paper. Her paper concept can now be deployed in new types of explanations and predictions: ones involving pen and ink. Does the child's old concept of paper necessarily change in virtue of this, simply because the range of explanations and predictions she can make with the paper concept has expanded? Surely not.

I have been pointing out counterintuitive consequences of the TT approach in virtue of its association with meaning holism. But now let me return to the core thesis of developmental TT: that the incidence of mentalistic conceptual change in childhood—assuming it can be established—would prove the occurrence of theoretical change. This contention is also problematic, because additional arguments would be needed to demonstrate theoretical change. Not all *conceptual* changes are *theoretical* changes. A person might change his concept of a "message," for example, from a communication written on a slip of paper to a communication conveyed by any kind of medium. This would be a change in how a term is conceptualized, but not obviously a change in theory. Of course, much depends here on what is meant by *theory*. But if a theory is construed as a set of law-like generalizations (a common construal by developmental theory-theorists), then not all concepts are theoretical concepts. So conceptual change per se does not suffice to establish theoretical change.

Proponents of TT often argue that ST is unable to explain the chronology of mentalization in young children. For example, can ST account for the fact that the desire concept is acquired earlier than the belief concept? Bartsch and Wellman (1995a, pp. 178–179) claim that ST has no satisfactory account of this. One possible answer is suggested by Harris (1996): that there is a social premium on the communication of desire as compared with the communication of belief. A great deal of child-adult activity involves collaborative action, the success of which depends on the communication of the desires of each. This premium on desires, with beliefs left unspoken, might explain the child's greater preoccupation with desire and consequent earlier mastery of the language of desire. Another possible explanation of the early emergence of desire talk is a more salient phenomenology. Desires may recruit comparatively greater attention, directed at finding means to their fulfillment. The phenomenological explanation is

admittedly tenuous, because, as Bartsch and Wellman indicate (1995a, pp. 180–181), the phenomenology of knowledge also has salient dimensions.

But what is TT's *own* explanation for the comparatively early emergence of desire talk? Bartsch and Wellman's explanation is that the desire concept (at least the 2-year-old's desire concept) is nonrepresentational, and therefore easier to grasp than belief. This explanation works as long as the following two claims hold: (1) The (early) desire concept is nonrepresentational. (2) The belief concept is (always) representational. But both of these assumptions are open to challenge. My previous discussion challenges claim 1. Claim 2 is challenged by other theory-theorists, who deny that belief is always understood representationally. So the alleged asymmetry in representational status between desire and belief is tenuous.

In contrast with developmentalists, philosophers usually regard beliefs and desires as symmetrical in representational status, for reasons similar to those reviewed above. To make a convincing case for a representation-based explanation of the earlier emergence of desire talk, TT proponents need more compelling evidence of representational asymmetry. Now suppose, for the sake of argument, that solid evidence for such an asymmetry were forthcoming. Why couldn't ST then invoke this same asymmetry to explain the earlier emergence of desire talk? It is not clear why such an asymmetry would not be usable by ST. Thus, I see no reason why ST is at a comparative disadvantage in trying to explain the chronology of mentalization.

### First-Person Attribution of Mental States

People detect and report mental states of self and of others. How is it done in each type of case? Do they use the same processes and heuristics for self-attribution as for other-attribution? Or are the processes fundamentally different?

The standard TT view, clearly enunciated by Gopnik (1993), is that the processes are exactly the same. We adults believe that our knowledge of our own mental states is substantially different from our knowledge of other people's mental states (much more "direct" and error-free), but this is just an illusion. To support this view, Gopnik cites evidence that 3-year-olds make certain consistent mistakes about their own immediately past mental states, and these mistakes parallel mistakes they make about others (Gopnik

and Astington 1988; Gopnik and Slaughter 1991; Wimmer and Hartl 1991). Gopnik takes this as evidence that we have no "privileged access" to our own mental states and that the processes used to detect our own states are the same as those used for others.

Defenders of privileged access maintain that it applies only to current mental states. Gopnik's evidence concerns past mental states—not long past, but nonetheless past. Moreover, in surveying the evidence Gopnik overlooks the fact that the very experimenters she cites, herself included, routinely rely on the *accuracy* of children's reports of their own current states. This point is well emphasized by Harris (1992, 1993). Gopnik and Slaughter (1991) showed 3-year-olds a familiar crayon box, asked what they thought was inside, then showed them the actual contents (candles), and finally asked them what they had earlier thought was inside. All 3-year-olds were accurate in response to the first question, about their then-current thoughts, at least as far as could be judged. The experimenters themselves assumed that the children answered this question accurately, an assumption on which they relied in concluding that the children's later reports were mistaken. So if we focus attention on *current* self-reports (the nub of the issue), the evidence does not favor TT. Instead, privileged access seems better supported.

In any case, Gopnik has told us next to nothing about exactly how one manages to detect one's own current states. Does one proceed by theoretical inference from observation of one's own behavior or of environmental circumstances that are potential causes of one's mental states? I have previously dramatized the implausibility of this approach by an example of waking up with a morning headache (Goldman 1993). One can immediately detect and classify such a headache without any previous information about its observable causes or effects. Here let me elaborate the point with a different example: a current intention. I now have an intention to snap my fingers in the next 10 seconds. How do I know that I have this intention? I have not detected any observable environmental cause of the intention, nor have I performed any overt behavior that might provide some clues. It is hard to resist the explanation that I have some sort of direct access to my intending state. Other advocates of TT, including Nichols and Stich (2000), are now coming around to this traditional and compelling approach. The details of privileged access are exceptionally controversial, and this is not

the place to address them. But so long as TT rejects all forms of privileged access, it will face serious difficulties.<sup>9</sup>

Critics of an introspectionist or direct detectionist approach sometimes suggest that it is incompatible with conceptual change. According to Gopnik and Wellman (1992), simulationism conjoined with introspectionism implies that "the child (or adult) doesn't need and doesn't appeal to a theory of mind, a conceptual understanding of mental states, to predict behavior or understand others. Instead she simply runs a perfect working model of a mind, her own mind. . . . Moreover, access to [her] own states requires no inference or interpretation, no conceptual intermediaries, no theorizing, you simply read them off." (ibid., pp. 159–160) Similarly, Bartsch and Wellman (1995a, p. 175) say this: "... when reasoning about minds [according to ST], we do not resort to concepts but rather more simply to our firsthand experience of mental life." These passages wrongly suggest that introspectionist ST excludes conceptual intermediaries and conceptual understanding. Introspectionist ST does not claim that concepts are unnecessary or dispensable in the tasks of detecting one's own states or detecting the states of others. The main point to emphasize is that detection or attribution of a mental state, whether in oneself or in another, is a conceptual act: it consists in classifying the target state as of one type or another. In the case of propositional-attitude states, the type is always complex, including both an attitude type and a specific content. The attributor must decide whether the target state is a belief, a desire, a wish, or a fear, and what its specific content is. All these classifications require concepts. Perhaps a type just is a concept. Of course, introspectionist ST may depart widely from TT in its specific account of mental types or concepts—in other words, in its account of how ordinary attributors conceptualize desires, beliefs, and fears. Introspectionists might wish to link these conceptualizations (partly or wholly) to phenomenological materials rather than to causal-functional relations (Goldman 1993), but this does not mean that proponents of introspectionist ST hold that concepts are irrelevant. Moreover, various types of conceptualizations may figure critically in the simulation heuristic involved in third-person attribution. For example, to execute a successful simulation of another person's thought one may have to "quarantine" or "overwrite" one's own knowledge. And it may be necessary to "flag" the simulated states of another person as simulated, pre-

tend, or make-believe (Harris 1991, 1995; Harris and Kavanaugh 1993). These activities require conceptualization of some sort, though not necessarily theoretical conceptualization. Children lacking the appropriate conceptualizations will not succeed at folk-psychologizing tasks.

### Simulation, Re-Creation, and Pretense

The fundamental idea of TT is that a mental-state attributor arrives at an attribution to a target person by reasoning theoretically from observed behavior of the target and/or from environmental conditions that might influence his or her mental state. Or, if the attributor has previous knowledge of some of the target's mental states, he or she might reason from these previously known mental states to others. The crucial point is that theoretical reasoning is the fundamental method or heuristic that attributors deploy.<sup>10</sup>

The prime example of theoretical reasoning, though not the only possible one, is reasoning in accordance with (believed) laws—specifically, *psychological* laws that relate mental states to one another and to inputs and outputs.

The process of simulation posited by ST stands in sharp contrast to this story. In using the simulation heuristic to detect a target's mental state (for example, a future decision), an attributor begins by taking the target's "perspective." That is, he *pretends* to be in certain states the target is in, as suggested by his previous information about the target (gleaned, perhaps, from previous simulations). He feeds these "pretend" starting states into an appropriate cognitive mechanism of his own (for example, a decision-making mechanism) and lets it operate on them. This mechanism outputs a new state (e.g., a decision), and the attributor then attributes that decision to the target. In other words, the attributor tries to make his own mind "emulate" the mental sequence the target will go through (or, in the case of attributing a past state, did go through). The heart of this procedure is that the attributor tries to *reproduce* or *match* what transpires in the target—something not featured in the TT approach at all. The process of mental reproduction or matching substitutes for theoretical reasoning in accordance with psychological laws, which are unnecessary under the simulation heuristic. This is not to say that simulation never involves reasoning. On the contrary, if an attributor simulates a process of practical reasoning, he

himself engages in practical reasoning. But it is not an instance of theoretical reasoning, which is what TT always requires. There are special cases in which simulation even involves theoretical reasoning, as when an attribute simulates a physicist, an engineer, or any other target person engaged in theoretical reasoning. A simulation of theoretical reasoning may even feature beliefs in laws of some sort. But when the target reasoning involves physical subject matter, the simulator will only appeal to (possibly "pretend") beliefs in *physical* laws, not to beliefs in *psychological* laws. Notice that ST need not claim that *all* mentalizing takes the form of simulation; it need only claim that mentalizing often takes that form, or that simulation lies at the roots of mentalizing.<sup>11</sup>

The popularity of TT as an account of folk psychology probably stems in part from the fact that the "dominant explanatory strategy" in cognitive science, as Stich and Nichols (1992) put it, is to view cognition as the acquisition and application of internally represented knowledge structures. These same authors regard ST as potentially "an insidious threat to traditional theories in cognitive science" (Nichols, Stich, Leslie, and Klein 1996, p. 40). However, I believe that the use of simulation in mentalizing tasks would not be an island in a sea of theorizing, and its postulation is certainly no threat to cognitive science. Simulation is an intensively deployed type of heuristic in many sectors of human cognition. Once this fact is appreciated, its plausibility as a hypothesis about folk psychology will have a better ring in the ears of cognitive scientists.

The first point to emphasize here is that human cognition is a large collection of heterogeneous mechanisms that evolution has found useful. The visual system alone incorporates an assortment of heuristics. Even depth perception, just one task of vision, is subserved by a bundle of disparate heuristics. So we should not be surprised to find an eclectic bag of tricks in nature's repertoire, rather than one homogeneous mode of cognizing. Furthermore, the device of mimicry or synchrony is present in a wide range of physical, physiological, and mental dimensions of human life. (For a review, see Levenson and Ruef 1997.) Young children and adults revel in mimicry, whether it involves postures, vocalizations, or facial expressions.

The generic trick I wish to highlight, however, has an epistemic, or information-gathering, theme. I shall call it *investigation through re-creation*.

Human beings engage in a variety of behaviors, both overt and purely mental, in which they try to "re-create" scenarios either previously observed or anticipated as possibilities. I conjecture that the function of such behaviors is to investigate, study, or explore the actual or possible scenarios that they model. In other words, to gain information about target scenarios, people create models that can disclose some of their properties. I do not say that these re-creative activities are always executed with this conscious aim. Rather, nature has given us propensities to engage in such re-creative activities because they have, as a frequent by-product, the disclosure or discovery of useful information. Mental simulation, I conjecture, is one specimen of this generic type of heuristic.

I start with direct behavioral imitation. Meltzoff and Moore (1977, 1983) discovered that newborn infants have an innate ability to imitate facial expressions. Imitation remains an important proclivity among older infants and young children (Meltzoff 1995; Meltzoff and Gopnik 1993). Clearly, direct behavioral imitation re-creates an observed scenario. Another trait of all normal children is pretend play, in either solo or cooperative variants. Pretend play is not direct imitation, but it is quasi-imitation of a genre of activity. When a child pretends to talk on a telephone by using a banana as a prop (Leslie 1987), he is not precisely imitating normal telephone usage, but the (overt) pretense does re-create selected aspects of the modeled activity. Similarly, the entire idiom of drama is a medium in which people imitate or impersonate individuals—typically not actual individuals or events but persons and scenarios that might exist or occur. The apparent cross-cultural universality of the dramatic idiom throughout recorded time strongly suggests an innate basis for this kind of activity.

I turn now to purely mental re-creation. Visual imaging is a form of mental re-creation. If I visually imagine the rooms in my house, I re-create (to some approximation) the visual perceptions associated with being in those rooms and looking around. Prospective re-creation occurs when one engages in planning activity. A person in the process of climbing a tree or a cliff may imagine how he would be positioned if, in his next move, he were to place his left foot onto a particular branch or ledge.

My general idea is that these forms of re-creation are valuable because they characteristically yield useful information. Visually imagining the rooms in one's house can yield information about the layout of the rooms,



or about how many windows are in the house. Imagining the result of a prospective climbing maneuver can yield information about the effectiveness or ineffectiveness, the riskiness or nonriskiness, of that move.<sup>12</sup> Attempts at behavioral imitation can yield other types of information. They can help the imitator learn what is required, in motor terms, to reproduce the target action. And they can help the actor discover the physical and social consequences associated with that type of behavior, including responses from other members of the actor's community. These are all valuable pieces of information, for children especially. Pretense has similar epistemic payoffs, at least in many cases. A child who practices talking on the telephone in "pretend" situations will be more skilled at doing the real thing and will make fewer mistakes when it "counts." She may learn the expected elements and proprieties that govern and facilitate telephonic conversation. Impersonation and drama help people learn how circumstances and situations different from their own will be experienced.

I do not claim that every act of behavioral imitation, pretense, or imagination yields useful informational bounty. Nor do I claim that the *only* benefits of re-creative acts are epistemic ones. My conjecture is that, in the main, these kinds of re-creative activities disclose useful information for creatures who need to chart a course through a physical and social environment. This is why these types of activity proved adaptive and why evolution wired us up with these propensities, though we make little or no conscious calculation of their epistemic benefits. Gopnik and Meltzoff (1997) make a convincing case for the thesis that infants and children are inquiry-driven creatures, a characteristic that surely has a genetic basis. I am suggesting that children's dispositions to engage in a variety of re-creative activities are specimens of this trait of inquiry-drivenness. My next suggestion, obviously, is that mental simulation has similar roots. If mental simulation enables us (often if not always) to make correct mental-state attributions to others, this epistemic payoff will be useful enough to have been favored by evolution.

### Mirror Neurons and the Simulation Theory

Is there more specific evidence for evolutionary origins of mental simulation? A plausible case can be made for this in the discovery of *mirror*

neurons in macaque monkeys. Certain neurons in the monkey's premotor cortex discharge not only when a monkey initiates specific goal-oriented acts but also when it observes another monkey (or a human) initiating the very same acts (Rizzolatti, Fadiga, Gallese, and Fogassi 1996). All mirror neurons discharge during specific goal-related motor acts, such as grasping, manipulating, or holding an object. The discharge of a specific group of mirror neurons is correlated with having a particular motor goal or intention. But the very same mirror neurons discharge when a monkey does not act but merely observes another monkey (or a human) engaged in that type of goal-related action. Vittorio Gallese and I conjecture that mirror neurons are a primitive version of, or a precursor to, mental simulation (Gallese and Goldman 1998), insofar as they are a vehicle by which an observer *mimics, resonates with, or re-creates* the mental life of a target. (In the observer's case, the motor plan is not executed. It seems to be taken "off line," which resembles the hypothesized pattern of mental simulation.)

There is evidence of a similar mirror system in humans. When people observe the actions of others, such as the grasping of an object, there is a marked increase in motor evoked potentials from the same hand muscles that are recruited when those actions are performed (Fadiga, Fogassi, Pavesi, and Rizzolatti 1995). Clinical evidence of such a phenomenon is found in so-called *imitation behavior* (Lhermitte, Pilon, and Serdaru 1986). Patients with prefrontal lesions compulsively imitate gestures or even complex actions performed in front of them by an experimenter. This behavior is explained as arising from an impairment of inhibitory control that normally governs motor plans. Apparently, when observing someone else perform an action, a normal human generates a plan to do the same action, or an image of doing it himself. This plan is inhibited—and thereby taken "off line," as simulationists like to say—so that it does not yield motor output; but such inhibition is impaired in the relevant patient population. This is further evidence of the mental "re-creation" of a target's action plans, of the very type ST postulates.

Nothing I have said here constitutes conclusive evidence for simulation as the core mechanism of interpersonal mind reading. Nonetheless, there is converging evidence of its inherent plausibility as such a mechanism. We have seen, moreover, that the vaunted advantages of TT and the putative



weaknesses of ST do not withstand scrutiny. Thus, ST deserves to be treated, by psychologists as well as philosophers, as a very serious contender for the key to the mystery of folk psychology.

### Acknowledgments

This chapter has benefited greatly from the extremely helpful commentary, both critical and constructive, of two of the volume's editors: Lou Moses and Bertram Malle. I also thank Henry Wellman and other conferees in Eugene for their patient attempts to help me understand things their way. I can only hope that my stubbornness in the pursuit of my own vision may ultimately have a community-wide epistemic payoff.

### Notes

1. Perner (1991, pp. 115–116) does not restrict the objects of desire to *actual* objects or states of affairs, but he does agree with Wellman in not importing *representations* into early conceptions of desire.
2. I introduce *world-related* to replace *connectionist*, which has a distracting, irrelevant association with parallel distributed processing.
3. The problem can be applied to cases involving intentions drawn from Meltzoff's (1995) study of 18-month-olds. Children saw an adult try and fail to execute some novel action. They proceeded to "re-enact" the action that the adult presumably intended. Meltzoff's findings strongly suggest that these children ascribed this intention to the adult. According to the world-related story, what is the token action that is conceived of as the relatum of the adult's intention? Could it be the *child's* own action? But the adult's intention is for *him* to do the action, not the child; so why would the child's performance be the relatum of the adult's intention?
4. I say "something like" because I don't wish to be committed, or to commit 2-year-olds, to a view of mental representations like, e.g., Fodor's (1981) language of thought. In fact, I am not happy with the putative exhaustiveness of the two conceptions of desire offered by Bartsch and Wellman: the world-related conception and the representational conception. However, since a full discussion of these issues would take us far afield, I restrict my discussion to these two options.
5. Thanks to Josef Perner for calling attention to the need to avoid this confusion.
6. In the past, when false-belief errors seemed to show that young children lack a representational model of belief, it was reasonably natural to infer that they would also lack a representational model of desire (at the same and earlier stages). But in the wake of the recent evidence about false-belief tasks, cited above, we must reassess earlier assumptions about young children's models of belief. The same should be done for their models of desire.

7. Schwitzgebel (1999) presents evidence that some theorists—though not Bartsch and Wellman—*have* confused these matters.

8. For general doubts about the tenability of meaning holism, see Fodor and LePore 1992. Moses and Chandler (1992) also appeal to holism to criticize some aspects of orthodox TT, but their point alleges that an asynchrony between an early understanding of desires and a later understanding of beliefs would be *nonsensical* under the TT approach. My criticism does not make quite so strong a claim. At the same time, I find it very puzzling how the 3- or 4-year-old is supposed to work with a theory of mental states in which *some* mental states (desires) have *real-world* relata whereas other mental states (beliefs) have *representational* relata. It is radically unclear how these two families of relata can readily be coordinated with one another.

9. It is standard among theory-theorists to regard intentional states as "unobservable." For example, Leslie (1994a, p. 211) writes: "How is the preschool child able to learn about mental states when these are unobservable, theoretical constructs? Or put another way: how is the young brain able to attend to mental states when they can be neither seen, heard nor felt?" On other occasions, theory-theorists take a softer position. In one place, for example, Gopnik and Wellman (1994, p. 290) say that their preference for TT over ST "does not mean that we completely reject a role for first person knowledge in the formation and application of a theory of mind. Our own first-person psychological experience provides a body of data that strongly informs and shapes our concepts of mind." It is not clear to me, however, how this last passage fits with Gopnik's (1993) explicit denial of privileged access. Perhaps Gopnik and Wellman mean that phenomenological experience plays a role in (mental) concept *formation* but offers no noninferential method for *classifying* one's own current states. If this interpretation is right, then privileged access is still denied.

10. Theory-theorists occasionally concede that simulation can play *some* role in mental-state determination. For example, Gopnik and Wellman (1994, p. 280) write: "... we do not wish to deny an important role for the simulation of others' states of mind. A theory of mind must confront the problem of inferring the mental states of others in particular cases. Several devices can be used to aid in this process, but imagining what your own mental states would be rates high among them. We contend, however, that such simulations are theory-driven." If this point were made systematically, it might constitute an important dimension of compromise with ST. However, the suggestion that simulation is "theory driven" presents simulation in a somewhat different light than orthodox ST proposes. Simulation theorists draw a distinction between "theory driven" and "process driven" simulation, and advance the hypothesis that *process driven* simulation lies at the core of third-person mind-reading (Goldman 1989, p. 173). Some erstwhile theory-theorists have recently adopted an even more qualified position than the quoted passage from Gopnik and Wellman suggests. For example, Perner (1996) describes himself as advocating a "simulation-theory mix."

11. ST's willingness to admit that not *all* mentalization involves simulation may be a partial compromise with TT, but it is not a capitulation or "collapse" into it. Mentalizers may construct generalizations or schemata based on earlier instances of

simulation, and then use those schemata for new acts of mentalizing without simulation (Goldman 1989, p. 176). But under this scenario, the roots of mentalizing are still anchored firmly in simulation.

12. The value of strategy testing has been offered by Currie (1995) as an evolutionary explanation of imagination and of (mental) pretense. Currie credits Dennett (1978a). Neither of these authors generalizes this sort of explanation, as I am doing, to all varieties of re-creative activity.

## On the Possibilities of Detecting Intentions Prior to Understanding Them

Daniel J. Povinelli

Most authors in this volume are united in the opinion that the ability to conceive of intentions is a fundamental and universal aspect of how our species thinks about the social world. The question of how and when human infants and children develop this ability, however, inspires a good deal less harmony.

What scientific techniques can be brought to bear on this question? There are many, of course, but perhaps the most interesting—and most controversial—involves use of visual habituation-dishabituation procedures and related techniques. In such research, infants are allowed to repeatedly observe a stimulus that embodies some conceptual principle (e.g., that unsupported objects fall) until their visual attention to the stimulus wanes. Generally, it is assumed that this decline in visual attention signals infant's habituation to both their own visual experience of the stimulus and whatever conceptual interpretation of the stimulus they may possess. Once habituated (as measured by their viewing time), the infants are typically divided into an experimental group and a control group. These groups observe slightly different stimuli. The experimental group is typically presented with a stimulus that violates the conceptual principle embodied in the original stimulus but, significantly, is designed to be visually quite similar to the original stimulus. The control group is presented with a stimulus that is perceptually different from the original stimulus but which embodies the same conceptual principle. If the experimental group exhibits a greater rebound in visual attention than the control group, many researchers feel comfortable arguing that the infants are sensitive to the particular concept instantiated by the original stimulus. Although the general reasoning behind such procedures (offering pre-linguistic infants the opportunity to express what