

### 3 Intentional Agents Like Myself

Robert M. Gordon

#### 3.1 Introduction

According to Meltzoff, Tomasello, and Gallese, certain human responses to conspecifics have the following property: although they do not require possession of mental concepts, they nonetheless manifest an implicit "like me" recognition, a recognition of conspecifics as intentional or goal-directed agents like oneself. This is an important idea, one that I think is crucial to understanding how we can bootstrap ourselves into an explicit folk psychology. I don't think it has been developed adequately, however. Meltzoff, I believe, was the originator, and here I will try to point up some inadequacies in the way he conceives this "like me" recognition, namely, in terms of analogical inference. Then I will sketch a very different account, which I think is particularly consonant with some remarks of Gallese's. Lest it appear that I am pitting Gallese against Meltzoff, or indeed myself against Meltzoff, I should note that some of Meltzoff's writings (e.g., Meltzoff, 1995) seem to me quite congenial to the view I am presenting.

I should make it clear that I am not talking about how we "read" other minds or anticipate the behavior of mind-endowed entities. I am concerned only with how, without prior possession of mental concepts, we can implicitly recognize certain entities as intentional agents like ourselves.

#### 3.2 Constitutive versus Imitative Mirroring

Gallese (vol. 1, ch. 3) and Meltzoff (vol. 2, ch. 1) are each concerned with phenomena that fall under the category of *mirroring responses*: roughly speaking, responses brought about by *b*'s perception of *a*, in which *b* comes to have property *p* because *a* has property *p*. For example, because *a* does or did something (of a given description *d*), *b* does "the same" (that is,

something fitting description *d*): because *a* activates and executes a given motor plan, *b* activates "the same" motor plan; or, because *a* undergoes certain visceral responses (specifically, those characteristic of the emotion expressed on *a*'s face), *b* undergoes "the same" visceral responses (see Adolphs et al., 2000).

It is important to note that whereas Meltzoff is speaking primarily of the *imitative* mirroring of another's behavior, Gallese's discussion is more concerned with mirroring that constitutes part of one's very *representation* of the other's behavior (as explained in the next paragraph). This difference is crucial. If I try to imitate your behavior, I try to copy or match something I have perceived you to be doing, perhaps along with the manner in which you did it. However, for any actions for which I have the corresponding mirror neurons, in perceiving the behavior that I am now trying to match, my brain was already making use of a copying or matching procedure. As I observed you, one or more of my premotor neurons responded as if it were I who was carrying out the behavior. Now, as I imitate you (at least, if I do so successfully), presumably the same neurons that had previously responded *as if* I were carrying out the behavior will be activated again as I actually carry it out. The first response I will call *constitutive mirroring*, in that it was a constitutive part of my representation of your behavior; the second, I will call *imitative mirroring*.

According to the results cited by Gallese, the sight of other (living) human or humanlike bodies deposits in one's brain, not just a visual representation of their behavior, but also internal replicas of, among other things, the motor plans and visceral responses—and possibly even the lower-level intentions—that lie behind the behavior. Although these replicas may be implemented within my brain when I observe your behavior, that does not make them *my* intentions, urges, and motor plans. First, they are not *endogenous*. They are not produced by my own decision-making and emotion-forming processes. Rather, they are *exogenous* states, induced "from the outside" by observation of another's behavior. Second, thanks to processes that are usually automatic and often unconscious, these responses are mapped onto another human or humanlike body, ideally the one whose motor behavior or facial expression elicited the response. For example, I see my son's leg poised to kick a soccer ball, and my own leg involuntarily prepares to kick, but in a way that helps me to anticipate his kick, not my own, and also to recognize it as a kick toward his left, not toward my left. Even though this projection onto my son may emerge into consciousness, it is surely not something I have brought about by ana-

logical reasoning. I do not begin with a belief that something is going on in me, as opposed to in him, and then conclude with a belief that something is going on in him, as opposed to in me. In order for my mirroring to assist me in anticipating my son's kick, I needn't even be aware of my own leg's preparing to kick. And I don't theorize that my son must be intending and preparing to kick. Rather than infer from some intention of my own that my son has a certain intention, I find myself "getting behind" his behavior, as if it were my own.<sup>1</sup>

This phenomenology of "getting behind" is probably the cumulative result of a number of factors. The mirrored motor plan enables me to anticipate what his body will do, within his egocentric space (a kick toward his left), and, equally important, within his explanatory context: I am a defender, the goal I am defending is on the right, and I need to get the ball to the side without crossing it in front of the goal. It is these factors, as well as the resulting phenomenology, that justify calling the mirroring of his motor plans and behavior constitutive of my representation of his behavior.

(Concerning the relationship between constitutive and imitative mirroring, I will offer a hypothesis that is not essential to my argument but may be worth investigating. It seems reasonable to speculate that when I later recall your behavior with the purpose of imitating it, I reactivate not only a visual image but also the pattern of premotor and motor activation that occurred when I first observed your behavior. Then my actual, or overt, imitation will consist in, or at least build on, the now disinhibited reactivation of that pattern. Thus, when I imitate, I do not have to go back to a purely visual memory and then do a crossmodal mapping from visual to motor representations, for I have already captured your action in motor memory. I need only retrieve the pattern from memory and, as I suggested, reactivate it—this time, actually carrying it out rather than inhibiting it from overt expression.)

1. Buccino et al. (2001) establish that the mirror system in humans extends to perceived actions of the foot as well as of the hand and mouth. Beyond the mere replication of motor plans, when we observe object-directed foot actions such as ball-kicking, we engage parietal systems that are probably conducting higher-level analyses of the action. (I thank Vittorio Gallese for the reference.) Strictly speaking, the neuroscientific evidence does not yet show the replication of intentions. However, the phenomenology, as well as some of the research by Wolfgang Prinz (vol. 1, ch. 5), suggests that I replicated my son's *intention* to kick the ball to the left.

### 3.3 Meltzoff on the Analogy of Self and Other

In volume 2, chapter 1,<sup>2</sup> Meltzoff writes:

Human acts are especially relevant to infants because they look like the infant feels himself to be and because they are events that infants can intend. When a newborn sees a human act, it may be meaningful: "That seen event is like this felt event." (p. 74)

Thus in Meltzoff's view, the infant uses an argument from analogy of the form: When I produce behavior of type *x*, I feel a certain way *f*; therefore, when a similar body does *x*, the behavior was probably produced by another subject—another "I"—that feels the same way *f*.

According to Meltzoff, such an inferential process is well within the capacity of the human infant. However, the capacity for analogical reasoning is not the only concern. To apply the argument would require the following additional capabilities:

1. being able to identify one's own behavior in a way that allows comparison with the observed behavior of another body and
2. being able to identify one's own feeling or experience *as such* (i.e., interpret it as something that is going "within me," in the appropriate sense; that is, subjectively, as opposed to "out there in the world" or in someone else).

The first capability would be particularly problematic in the imitation of facial expressions, since the infant has no visual perception of its own current facial expression. Even adults have difficulty (I do, in any case) associating their own current facial configuration with a visual image. Therefore, I do not think Meltzoff can be right in asserting, as he does in chapter 1 (vol. 2), that "when infants see others acting similarly to how they have acted in the past, they project onto others the mental experience that regularly goes with that behavior. This could not occur if infants saw no equivalence between their acts and those of others" (p. 57).

More generally, infants would have trouble with the second capability, because it would demand considerable conceptual sophistication to under-

2. Meltzoff has defended the analogical inference account in numerous other publications, including Meltzoff & Gopnik (1993). In personal correspondence, however, he notes that he did not mean that the baby "thought through a step-by-step formal analogy." Rather, as he states later (Meltzoff, 2002a, p. 35), because infants are able to "recognize the similarities between their own acts and those of others," the acts of others are imbued with "felt meaning."

stand that "this"—whether it be a particular sensation of pain or the phenomenal aspect of an action such as sticking out one's tongue—is just something that is going on within me, in the appropriate sense, that is, subjectively, as opposed to out there in the world or in another. Both of these capabilities would be required to make sense of the premise, "This is what is going on within me when my body is doing *that*," and thus to get an argument from analogy started.

For a further illustration of the problem, consider another, better-known neonatal tendency to mirror another's behavior: responsive crying. Infants, even neonates, exhibit emotional distress when they cry in response to the crying of other infants. To get an argument from analogy started, the infant would have to conceptualize as follows: "This distress (namely, the distress that I am 'directly' aware of) lies behind *this* crying, but it is not what lies behind *that other* crying I hear." But does the neonate, does even the older infant, have the sophistication to think herself into such a posture? I think not. What should be problematic for the infant is not *assimilation* (whatever is doing that crying must be undergoing what I am undergoing), but *differentiation* (whatever is doing that crying is something distinct from me).<sup>3</sup> Without the ability to differentiate between *a* and *b*, of course, there can be no analogical inference from *a* to *b*. At the same time, there would be no need for an analogical inference before the infant has begun to individuate minds and to think, "My mental state, *my* distress, is not what lies behind *that* crying."

An analogical argument may sometimes be applicable to "mature" imitative mirroring. Arguably, when I imitate your behavior, I may somehow take note of the inner states, such as intentions, urges, and perhaps even motor plans insofar as I am aware of them, that underlie my behavior. (Meltzoff speaks only of "feelings," but that seems an unnecessary limitation.) In imitating what you are doing with a box, I may find myself having the intention, say, to open the box. Then I might speculate, "Something like *this* may have transpired in you when you opened the box."<sup>4</sup> However, even if such an account sometimes applies to mature imitative mirroring, it certainly does not apply to the constitutive mirroring that Gallesse is concerned with.

3. For a relevant discussion, see and compare Gallesse (vol. 1, ch. 3) on the shared manifold, ED.

4. For the purposes of analogical argument, I would have to disregard some of my intentions. My intention to be imitating you, for example, would be an intention I should not project onto you.

### 3.4 Constitutive Mirroring and Intentional Explanation

In chapter 3 of volume 1 Gallese emphasizes that "we do not just perceive ... someone to be, broadly speaking, similar to us. We are *implicitly* aware of this similarity because we literally *embody* it" (vol. 1, ch. 3, p. 104, emphasis added). Later, elaborating on the relevant notion of embodiment, he cites Merleau-Ponty: "It is as if the other person's intention inhabited my body and mine his" (1945; English translation 1962, p. 185).

Gallese's discussion of embodiment (and Merleau-Ponty's discussion of habitation) seems to point toward something quite different than an argument from analogy; different, indeed, from any argument at all. His discussion of a shared manifold of intersubjectivity is suggestive, but I will offer what I think is a clearer picture of the way embodiment—in *contrast* to inference, whether analogical or not—might yield an implicit recognition that one's conspecifics are intentional or goal-directed agents like oneself.

For the kind of recognition I have in mind, what is necessary and sufficient is this: that I interpret their behavior under the same scheme that makes my own behavior, along with the intentions, motor plans, and visceral feelings that underlie it, intelligible to me; namely, the intentional scheme of reasons, purposes, and object-directedness. In the case of my endogenous visceral feelings, the brain typically incorporates them automatically into the "emotional coloration" of the eliciting object. Thus, when I gaze at the Grand Canyon beneath me, a large part of its emotional quality evidently comes from my sensory pickup of what is happening in my body. Presumably the brain selects the particular object to which the feelings are to be referred by consulting the emotion-formation system that produced the visceral response in the first place. In the case of endogenous intentions and motor plans, the brain evidently has ways of making their consequences unsurprising to us, probably by using effference copies and forward models. However, it also has ways of making the intentions and motor plans themselves unsurprising, by embedding them within a structure of reasons and purposes; I am running because it is raining, and doing so in order to avoid getting drenched. It seems a plausible hypothesis that these determinations too would generally be made by consulting the same system that produced the decision to run in the first place.

(A brief note on "consulting" the system. I do not mean to refer to a mysterious process of introspection, but rather to a hypothetical mechanism like one of the following: a hypothetical neural capacity to do a "trace" of the pathways and processes that led to a particular outcome, which is of course something we can set an ordinary classical computer to

do, except that a neural system would also assess "weights" at various nodes; a hypothetical capacity of decision-making and emotion-formation systems to conduct "what if?" experiments on themselves. For example, a system might subtract a particular input and see if that would make a difference in outcome. The latter hypothesis seems to me to fit with forward models of various kinds, and also with the way we deal consciously with counterfactual questions of the sort, "What would you have done if ...?" Generally, we seem to answer such questions by *deciding* what to do. See Gordon, 2002.)

The thesis that draws inspiration from Gallese's discussion of embodiment is this: The brain treats the exogenous replicas of another's motor plans and visceral responses in the same way it treats their like-coded endogenous counterparts. It seeks to make them unsurprising, to make sense of them, by fitting them to the "intentional" scheme of reasons, purposes, and object-directedness. It cannot do so directly, however, because it does not have access to the system that originally motivated them. Instead, it may "attempt," in one way or another, to produce in itself a like-coded *endogenous* response, one that matches the exogenous response it seeks to make unsurprising. Because it does have access to the system that motivates the endogenous response, the brain is able to consult it in assigning an intentional interpretation. Then it might assign the same interpretation, at least tentatively, defeasibly, to the matching exogenous response: a process of analysis by synthesis (see Kinsbourne, vol. 2, ch. 7, p. 168ff).

The general idea—a speculative idea, of course—is that, when a motor plan is induced exogenously, the brain will test various hypothetical ways of embedding it in an intentional scheme of reasons and purposes. This hypothesis testing (or hypothetico-practical reasoning, as I called it in Gordon, 1986) would engage productive processes such as practical reasoning, emotion formation, and decision making. In many cases, the hypotheses would be generated by a search of the shared environment.

I have already set out one instance of this. When I mirror my son's kick to the left, I also supply an explanatory context: I am a defender, the goal I am defending is on the right, and I need to get the ball to the side without crossing it in front of the goal. Within this context, his behavior makes sense; because I take it for granted that he is a smart player, it is even unsurprising. This is a complex case, however. The processing that would make his behavior unsurprising is complicated. It would be more illuminating to discuss a very easy case in which the brain might make sense of

an exogenous motor plan by fitting it to the intentional scheme of reasons and purposes.

You see your colleague reaching out and picking up an object. What you observe fits up your mirror neurons; if it did not, your brain might interpret the motion as it does other observed physical phenomena, calling on a theory or model. Thanks to your mirror neurons, however, visual perception deposits in your premotor cortex the motor plan for reaching out and picking up the same object. However, unlike motor plans that are produced in the normal way by your own decision-making system, this one arrives unmotivated, without reason or purpose. The object you find yourself picking up—or not quite picking up, stopping short of it—is a telephone. Specifically, it is your colleague's office phone.

Suppose that in fact your colleague is picking up the phone because it is ringing. What she is doing is answering the phone. Most of us, hearing a phone ringing nearby, might by habit have an initial impulse to pick it up ourselves. Then memory kicks in and inhibits the impulse. That is, we do not actually reach for the phone, because the phone that is ringing is not ours but another's, and since the other is nearby (and not incapacitated) it would be inappropriate for us to pick it up. Suppose that, as is plausible, your premotor cortex is activated: a motor plan is initiated for picking up an object with the size and location of the phone receiver, and perhaps also for bringing it to your ear. (If one's motor cortex were not activated in this way, I doubt one would report "an impulse to pick up the phone.")

In this case, it is likely that you would be independently activating the very same mirror neurons that were activated by observing your colleague. There would be two pathways to activation, one exogenous and one endogenous. Even though you do not have access to the systems that generated the exogenous, or incoming, motor plan, you do have access to those that are generating the endogenous, or outgoing, motor plan. This access allows you to give reasons and purposes for performing the action you would have performed, had it not been inhibited. It is obvious to you, for example, that you would not have been inclined to pick up the phone just now if it were not ringing, and that your sole purpose was to answer the phone. Then you would have a ready answer if asked why your colleague reached over to pick up that object. Not, of course, an infallibly correct answer, but a good first approximation, an answer likely enough to be correct that it could serve as a default answer.

Now consider an easy case in which the brain, by producing an endogenous counterpart, might make sense of a facial expression the sight of which induces an exogenous visceral response, representing the expres-

sion as directed toward an object. Suppose I am looking at someone whose facial expression induces an exogenous visceral response in me. My brain maps the response onto her face, thereby isolating it to some degree from my endogenous visceral responses: These feelings are hers, not my own. But, as with my own emotion-induced visceral responses, my brain looks for something in the world to which the response is to be referred. Typically, I find myself following the other's direction of gaze, halting at something she is obviously looking at. If the scene is complex, my gaze halts at whatever in her line of gaze *endogenously* produces in me the same or a similar visceral response that her expression is exogenously producing. If, for example, her face shows fright, my gaze halts at something frightening, something that induces in me, at least to a small degree, the visceral disturbances characteristic of fright. Sometimes the search fails to yield such an "objective correlative," and that is where imaginative transformation may come into play—often, an involuntary fleeting transformation, such as one's adoption of a child's perspective, from which, for example, what is not terrifying appears terrifying, or the converse. Sometimes such a transformation will succeed in yielding an endogenous match to the exogenous response induced by the expression. And sometimes not.

Each of these simple cases begins with something I assume the brain would find problematic: a visceral response, motor plan, or intention that is thrust upon it unmotivated. More precisely, the original of which it is a copy was motivated in a decision-making system and emotion-formation system other than its own, as if the brain were "possessed" by alien spirits. To avoid conflict with its endogenous productions, it maps the exogenous response onto an appropriate body. Exogenous plans and feelings need not integrate with those produced by one's own decision-making and emotion-forming processes. Rather, in effect, they will have been separated into distinct "I's," typically one per enduring human body.<sup>5</sup> Not only does it make sense of the behavior of another body to regard it as the expression of an inner mental life, it also makes sense of one's own inner mental life to assign a portion of it to the other body. For it avoids the disunity that would result if one had to "own" every stray motor plan, urge, and feeling that was injected exogenously into one's brain. What the brain does in these cases is, in a manner of speaking, to multiply the first person, so that exogenous plans and feelings are on the one hand assigned to a multiplicity of other bodies and on the other hand interpreted under the same

5. Only by way of this partitioning can one come to understand "I" as a true indexical, referring to one "I" or "self" among possible others; see Millikan (1993).



intentional scheme as their endogenous, truly first-person counterparts. This, according to my account, is what it is to implicitly recognize others as intentional beings like oneself.

I further speculated that in lieu of access to the systems that activate our exogenous responses, the brain might substitute a procedure of analysis by synthesis, producing a similarly coded endogenous response that it *can* analyze. Often, much of the work would be done by our common environment, together with our common biology and our socialization. For example, I respond as you do to the ringing of a telephone, and the same motor plan is independently activated endogenously as well as exogenously. Or I look at something, I see you looking at the same thing, and I get the same visceral response endogenously from the object you are looking at as I do exogenously from looking at your face. However, sometimes, as noted earlier, the exogenous activation initiates a search of the environment that halts when the same visceral response is produced endogenously. Sometimes an imaginative transformation is required for a matching endogenous response; and sometimes nothing does the trick. One way or another, the brain seems to be seeking an endogenous match to the exogenous intruder. Even the process of "getting behind" my son's kick to the left may involve, not only exogenous kicking, but also its endogenous replication. Not only do I automatically make the spatial shift that allows me to interpret my incipient kick to the left as a kick to his left; also automatically, I judge what to do in his "place" (i.e., in the role of a defender so situated) and proceed to do it—in an inhibited sort of way. If what I "do" endogenously is the same as what I "do" exogenously, then I shout, "Good move!" If not, then, perhaps, I criticize later. Aside from the considerable oversimplification, I think it a plausible speculation that something along these lines—the congruence or incongruence of exogenous and endogenous activation—may underlie some aspects of acculturation, such as instruction in a physical task.

### 3.5 What I Have Been Trying to Show

I have been trying to show how constitutive mirroring responses may manifest an implicit recognition of conspecifics as intentional or goal-directed agents like oneself, without requiring possession of mental concepts. My negative claim is that this implicit recognition is not the conclusion of an inferential leap from self to other. An analogical inference would begin with a premise concerning the states underlying my own behavior; more

particularly, those states of which I am aware. However, the mirroring phenomena I have been discussing are not "my own" in the requisite sense. If I am aware of them at all, I am aware of them as underlying the other's behavior, not my own. My positive claim might be put this way. The implicit recognition of conspecifics as intentional agents like oneself is a case of procedural rather than declarative knowledge. Specifically, the human brain will in fact seek the reasons and purposes behind the exogenous motor plan or intention, or the object to which the exogenous visceral feeling refers, just as it would for its own endogenous productions. If the brain does this, then it is treating the corresponding behavior, that is, the behavior that induced the exogenous response, as the behavior of an intentional agent.

I suggested at the outset that this implicit recognition is crucial to understanding how we can bootstrap ourselves into an explicit folk psychology. Bootstrapping is possible because intentional explanations in terms of reasons, purposes, and objects are at least implicitly mental. Even though there is no explicit mention of beliefs in, "I am running because it is raining," or of desires in, "I am running in order to avoid getting drenched," nonetheless these explanations, understood as intentional explanations, are true only if the corresponding mental state ascriptions and explanations are true. If I am indeed running because it is raining, that is, for the reason that it is raining, then I am running because I believe it is raining. And if I am running in order to avoid getting drenched, then I am running because I want not to get drenched. I am fairly confident that one of the principal avenues by which children come to develop the concepts of belief and desire is through the capacity to give such implicitly mental explanations of others' actions as well as their own. It would take several pages to set out how the ability to give these explanations can be parlayed into making explicitly mental (because I believe, because I want) explanations, but at least the seeds of such an account may be found in what I have written about ascent routines (Gordon, 1995b, 1996, 2000).

Mirroring systems probably play a very important role in "mind reading" by simulation (see Gallese & Goldman, 1998). If this is so, then analysis by synthesis may be the way, or at least a way, in which constitutive mirroring plays this role, making up for the fact that the brain lacks access to the systems that produced the responses it is mirroring. However, my main concern here has not been with whether and how constitutive mirroring might contribute to mind reading. What I have tried to show is how the human brain, by forcing exogenous responses into the same intentional

scheme that makes our endogenous responses intelligible to ourselves, implicitly recognizes the external sources of these responses as "intentional agents like oneself."<sup>6</sup>

### Acknowledgments

The author thanks Susan Hurley, Shaun Gallagher, Vittorio Gallese, and Natika Newton for helpful comments on earlier drafts.

6. For discussion relevant to this chapter, see especially Gallese (vol. 1, ch. 3) and Hurley (vol. 1, ch. 7). ED.

## 4 No Compelling Evidence to Dispute Piaget's Timetable of the Development of Representational Imitation in Infancy

Moshe Anisfeld

### 4.1 Introduction

Recent experimental work on imitation in infancy has challenged Piaget's theory and timetable (Piaget, 1951/1962, Part 1). Two aspects of Piaget's work have been criticized: his contention that imitation of invisible gestures (i.e., gestures the imitator cannot see when he or she performs them) could not occur until the third quarter of the first year, and his contention that deferred imitation of novel sequences of actions could not occur until the beginning of the second year.

The critics have marshalled empirical research that they interpret as showing invisible imitation in the neonatal period and deferred imitation at 6–9 months. This chapter argues that in both areas the empirical criticism of Piaget is not well founded. It removes a source of support for theories that attribute mental representation to young infants. In turn, it provides support for Piagetian theories that see mental representation as evolving gradually in the course of the first year.

The chapter starts with a brief summary of Piaget's theory to provide a context for his work on imitation. This summary is followed by an examination of the work on invisible imitation and deferred imitation.

### 4.2 Piaget's Theory of the Development of Representation

According to Piaget, (1951/1962, 1952/1963, 1954), in the first 6 months infants' functioning is nonrepresentational. The memories that young infants form of the stimuli they encounter are strictly tied to the sensory impressions of the stimuli and the motor adjustments that they elicit. These sensorimotor memories are elicitable by the stimuli that produced them; they are not otherwise available; that is, they are not represented independently.