

Perspectives on Imitation: From Mirror Neurons to Memes
Hurley, S. & Chater, N., eds. (MIT Press, 2005)

Chapter 15.

Intentional Agents Like Myself ¹

Robert M. Gordon

Abstract: Meltzoff introduced the important idea that an infant implicitly recognizes other human beings as “intentional agents like myself.” Such recognition is a step toward the explicit ascription of mental states. However, this “like me” recognition is not due, as Meltzoff sometimes suggests, to analogical inference, but to mechanisms that force one to interpret the behavior of others as if it were one’s own. Such an account is consonant with some remarks of Gallese’s. To defend the alternative account, a distinction must be made between two kinds of mirroring response, imitative mirroring and mirroring that is constitutive of the brain’s very representation of the other’s behavior. In the latter, motor plans, visceral responses, and other “first person” phenomena are imposed on the brain without motivation. The brain seeks to explain them as it does their internally produced counterparts, under the “intentional” scheme of reasons, purposes, and object-directedness. This, according to

the alternative account presented, is what it is to implicitly recognize others as intentional beings like oneself.

15.1. 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 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 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 will be 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.”

15.2. Constitutive vs. imitative mirroring. Gallese (2004) and Meltzoff (this volume) are each concerned with phenomena that fall under the category of *mirroring responses*: roughly, 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 of 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 notice 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 were 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 out the behavior. 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 human-like 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. For, 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 human-like body, ideally the one whose motor behavior or facial expression elicited the response. For example, I see my son's leg poised to kick the 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 analogical reasoning. I don't 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.² 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'm a defender, the goal I'm 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 don't 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.)

15. 3. Meltzoff on the analogy of self and other. In a recent paper,³ Meltzoff writes:

Human acts are especially relevant to infants because they look like the infant feels himself to be and because they are events infants can intend. When a human act is shown to a newborn, it may provide a recognition experience, “Lo! Something familiar! That seen event is like this felt event.” (Meltzoff, 2002a)

Thus, 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 . [*my paraphrase*]

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 worry. To apply the argument would require the following additional capabilities:

- (1) to identify one’s own behavior in a way that allows comparison with the observed behavior of another body
- (2) 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 (1) would be particularly problematic in the imitation of facial expressions, as the infant has no visual perception of its own current facial expression. Even adults have difficulty (I do, in any case) associating own their current facial configuration with a visual image. Therefore, I do not think Meltzoff can be right in asserting that

When infants see others acting similarly to how they have acted in the past, they project to others the mental experience that regularly goes with that behavior. This projection would not get off the ground if infants saw no equivalence between their acts and others' (Meltzoff, this volume).

More generally, infants would have trouble with capability (2), as it would demand considerable conceptual sophistication to understand that “this” – whether it be a particular pain sensation or the phenomenological 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).⁴ 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, 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."⁵ However, even if such an account sometimes applies to mature imitative mirroring, it certainly does not apply to the constitutive mirroring Gallese is concerned with.

15.4. Constitutive mirroring and intentional explanation. Gallese emphasizes that we do not just “perceive” that someone to be, broadly speaking, similar to us. We are *implicitly* aware of this similarity, because we literally *embody* it. (Gallese, 2004, 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.

Gallese's discussion of embodiment (and Merleau-Ponty's of habitation) seems to point toward something quite different from 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 just this: that I interpret their behavior under the same scheme that makes my own behavior, along with the intentions, motor plans, and gut feelings that underlie it, intelligible to me: namely, the “intentional” scheme of *reasons*, *purposes*, and *object-directedness*. In the case of my endogenous gut 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 picks out 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 efference 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, as 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, this volume, p. 000).

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'm a defender, the goal I'm 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 makes 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 stirs up your mirror neurons; if it didn't, 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. To keep the story simple, let's suppose that in fact the phone is ringing. You have no reason to pick up your colleague's phone when it rings, especially if she is present. But your hand starts to reach for the phone, until the memory that the phone is your colleague's kicks in and inhibits execution of the plan. Here you would be *independently activating the very same mirror neurons* that were activated by observing your colleague. What is important about this truncated action of yours is that, even though the movement of your arm and hand toward the phone has been inhibited, you have in effect *motivated*, that is, *provided a motivation for*, the exogenous motor plan. One might prefer to say that you have

motivated a *matching* plan. However, if the neural encoding of the plan is the same – the same neurons activated, now endogenously as well as exogenously – then we might as well call it *the same* plan. Now let's suppose your brain has a mechanism, such as those rough-sketched earlier, that can query the system that produced the decision to reach out for the object. The operation of this mechanism allows you to give *reasons* and *purposes* for so acting. It is obvious to you, for example, that you wouldn't be inclined to pick up the phone just now *if it weren't ringing*, and that your sole purpose was *to answer the phone*. (You were not preparing to *initiate* a phone call.) 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 in me an exogenous visceral response, representing the expression 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 needn’t 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.⁶ 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 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 motivated our exogenous responses, the brain might substitute a procedure of analysis by synthesis, producing a similarly coded endogenous response which 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 ring 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.

15.5. 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 isn't 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 gut 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,” nor 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 “mindreading” 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 mindreading. 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.”⁷

¹ The author thanks Vittorio Gallese and Natika Newton for helpful comments on earlier drafts.

² Buccino et al (2001) establishes 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, the neuroscientific evidence does not yet show the replication of intentions. However, the phenomenology, as well as some of the research of Wolfgang Prinz (2004), suggests that I replicated my son's *intention* to kick the ball to the left.

³ 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 in Meltzoff, 2002a, 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.”

⁴ Editors’ note: for relevant discussion, see and compare Gallese (2004) on the shared manifold.

⁵ 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.

⁶ 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.

⁷ Editors' note: for discussion relevant to this chapter, see especially Gallese (2004) and Hurley (2004).