*** Submitted version. Please do not cite without permission.*

CHAPTER X

MENTAL VS. EMBODIED MODELS OF MIRRORED SELF-RECOGNITION: SOME PRELIMINARY CONSIDERATIONS

JOHN L. SCHWENKLER,
DEPT. OF PHILOSOPHY, UNIV. OF CALIFORNIA, BERKELEY (U.S.A.)

A considerable body of recent work in developmental psychology and animal behavior has addressed the cognitive processes required to recognize oneself in a mirror. Most models of such "mirrored self-recognition" (MSR) treat it as the result of *inferential* processes drawing on the subject's possession of some sort of mature "self-awareness". The present chapter argues that such an approach to MSR is not obligatory, and suggests some empirical grounds for rejecting it. We also sketch the outlines of an alternative, "embodied" theory of MSR, and propose a way to evaluate it using the tools of adaptive robotics.

1. The received view

According to the "Received View" (RV) of mirrored self-recognition, MSR is the product of an *inferential* process drawing on the recognizer's possession of a particular sort of *self-awareness*. By reviewing the experiments behind the development of RV, we can unfold its logic.

1.1 the mark test

In a classic study (Gallup 1970), mirrors were introduced into the environments of solitary pre-adolescent chimpanzees. Over a period of several days, researchers observed a dramatic decrease in social responses directed to the mirror images, together with an increase in behaviors directed toward the self (e.g., grooming) with the aid of the mirror. In order to add more direct experimental support to the claim that these behaviors did indeed constitute self-

recognition, the researchers then anesthetized the chimpanzees and used red dye to mark their faces. Upon recovery from the anesthesia, the animals were reintroduced to the mirrored environments and directly observed for a period of thirty minutes to determine the number of times that the marked portions of their skin were touched. While the mirror-experienced animals used the mirrors to engage in a remarkably sophisticated range of mark-directed behaviors, control subjects with no prior experience with mirrors made no mark-directed responses at all when anesthetized, marked, and confronted with mirrors for the first time. Furthermore, the researchers observed that stump-tailed macaques given prolonged exposure to mirrors in a comparable situation and then tested in the same fashion exhibited no mark-directed behaviors at all, showed little decline in social behaviors directed towards their mirror images, and gave virtually no evidence of MSR.

The use of the "mark test", sometimes in combination with more informal observation of other possible self-directed behaviors, has since become common in research in animal behavior as well as developmental psychology.[1] The standard line of thought is that if an animal exhibits self-directed behavior toward a mark which is invisible without the mirror, then it must have recognized itself therein, and if it fails to behave in such ways, then it must have failed to recognize itself. Whether this methodology is acceptable, however, is not our present concern, since it is the cognitive mechanisms that have been proposed to *account for* MSR–whatever its behavioral evidence–that are in question here.

1.2 self-awareness

In his original paper, Gordon Gallup, Jr. concluded the following from the experiments we have just surveyed:

> Recognition of one's own reflection would seem to involve a rather advanced form of intellect; it is known, for example, that at least some mentally retarded children apparently do not have the capacity to recognize themselves in mirrors … Moreover, insofar as self-recognition of one's mirror image implies a concept of self, these data would seem to qualify as the first experimental demonstration of a self-concept in a subhuman form. (Gallup 1970, p. 87)

The idea that there is *some* sense in which MSR requires "a rather advanced form of intellect" and "implies a concept of self" is surely unproblematic, but it

[1] For the earliest use of the "mark test" on humans, see Amsterdam 1972.

is much more difficult to explain the exact sense in which these implications hold. For example, even if macaques are unable to recognize themselves in mirrors, there are many other ways in which they–like nearly all other "subhuman" animals–obviously *do* engage in behaviors that draw on something like self-recognition: for instance, using proprioceptive feedback to differentiate self-movement from changes in their environment, using vision to coordinate bodily movement, or even grooming the areas of their bodies that *are* immediately visible to them.[2] Most researchers, however, are after something quite specific when they talk about the involvement of "self-concepts" or "self-awareness" in MSR.

Gallup, for instance, writes that "*An organism is self-aware to the extent that it can be shown to be capable of becoming the object of its own attention*" (Gallup 1987, p. 3). This claim, however, is ambiguous in the same way as noted above: there is *some* sense in which almost *all* organisms can "become the objects of [their] own attention", for example if they directly perceive parts of their own bodies. What is really central, it seems, is that the animal in possession of a self-concept is supposed to be able to become aware of itself *as such*: that is to say, in attending to itself, it is *aware that it is itself to whom it is attending*. In addition, it seems essential that this awareness be what we might call "reflective": it is not enough that the animal simply *use* information about itself in certain ways (as in visually-guided reaching), but the animal must be *conscious* that this is what it is doing. Gallup's proposal is that MSR requires exactly this sort of potential for reflective self-awareness: it is only insofar as the animal is *reflectively* aware *that the animal it is observing is itself* that it uses the mirror image for grooming, exploration, and so forth.

1.3 the inferential model

According to the Received View, self-awareness figures in MSR in a very specific way: namely, within an *inference* that the self-recognizing organism must make. The idea here is that the recognition that it is *oneself* that one sees in the mirror is something that takes one beyond what is immediately given: as we have seen, chimpanzees initially regard their mirrored selves as conspecifics, and macaques seem unable to regard them as anything other than that. Similarly, since the capacity for MSR is something that must be *acquired*, it is supposed that the "learning curve" hinges on the organism's capacity for a bit of abstract reasoning, for example:

[2] See Gallup 1977: "While many organisms are ostensibly conscious of different features of themselves as the result of visual, tactile, chemical, and proprioceptive feedback, in principle this is quite different from self-consciousness" (p. 3). For more on the varieties of self-awareness, see the essays collected in Neisser 1993.

1. Mirrors usually reflect the objects directly in front of them.
2. I am standing directly in front of this mirror.
3. Therefore, the object I see in the mirror is (probably) me.

In this inferential progression, the conclusion (3) is an example of the "reflective self-awareness" discussed in §1.2, while there is a more minimal sort of self-awareness (i.e., an awareness of one's position in space) that figures in premise (2). Other versions of the inferential model, however, seem to have reflective self-awareness figuring immediately as a premise, for example:

> The unique feature of mirror-image stimulation is that the identity of the observer and his reflection are necessarily one and the same. The capacity to correctly *infer* the identity of the reflection must, therefore, *presuppose* an already existent identity on the part of the organism making this inference. (Gallup 1977, p. 334; emphasis added)

(It is admittedly unclear what Gallup means here by "already existent identity".) In any case, whatever the exact nature of such inferences[3], the key point to note is that according to RV they are the only way that MSR is possible.

2. Problems for the received view

Despite its widespread acceptance, there are some aspects of mirrored self-recognition that RV is not well-equipped to account for. In particular, a closer look at the trajectory along which human children acquire the capacity for MSR suggests that it may not be an inferential process at all, and the behavioral profiles of neurologically impaired subjects who have lost the capacity for MSR do not suggest that anything like inferential deficits are at play. In this section, we review this evidence and explain the challenge it poses to RV.

2.1 the learning curve

Researchers studying MSR are generally interested in its upshot: that is to say, they are interested in how animals *succeed* in recognizing themselves in mirrors, and in the conclusions that can be drawn from this about these animals' cognitive lives. What has been less exhaustively studied, however, is the trajectory along which this capacity is *acquired*. A recent paper by Mary

[3] For further discussion, see Mitchell 1993a, 1993b.

Courage and her colleagues (Courage et al 2004), however, investigates exactly this issue.

Unlike standard "latitudinal" analyses of the development of MSR, the Courage et al study investigates this phenomenon "longitudinally", by regularly testing children's responses to mirrors over a long period of time. In this way, they show that the capacity for MSR does not arise all at once when a child reaches a certain age. Instead, they observed children who would succeed in self-recognition one week and then fail for several weeks following before finally becoming consistent self-recognizers, or fail to pass the "mark test" in a given week even while they exhibited clear affective responses (such as embarrassment) to their altered reflections.

What is significant about this for our purposes is that this sort of piecemeal, on-again-off-again learning curve is not the sort of thing that easily suggests an *inferential* model of the progression. Intuitively, an inference is the sort of thing one simply makes or fails to make, and we are generally unwilling to regard someone as having made an inference in a certain situation when he or she fails to make that inference in a situation of exactly the same sort just a bit later. Additionally, the variability among the responses of children who *did* manage to self-recognize–again, involving affective, behavioral, and more clearly cognitive aspects–suggests that MSR involves a range of components that are not as straightforwardly "mentalistic" as RV makes them out to be.

2.2 the "mirror sign"

Another important body of evidence against RV comes from work in clinical psychology involving brain-damaged subjects whose capacity for MSR is impaired.[4] Such disorders are obviously fascinating in their own right, but the way they bear against RV is striking.

Recall that according to RV, mirrored self-recognition is the product of an *inferential* progression, perhaps combining knowledge of the state of one's own body with some general knowledge about the ways mirrors work. The natural way for an adherent of RV to explain these sorts of delusions, then, is in terms of either a generally impaired capacity for inferential reasoning of the sort sketched in §1.3, or a failure to believe one or more of the premises involved in such inferences. The difficulty, however, is that patients with the mirror sign tend to exhibit neither of these deficits: they need not be cognitively deficient in the way that the first explanation suggests (more precisely: if they are cognitively deficient, this is usually not the apparent cause of the mirror sign),

[4] For a recent review of the relevant literature, see Postal 2005.

nor are they generally lacking in explicit knowledge about themselves or the nature of mirrors.

Consider, for example, the following exchange between an examiner and a mirror sign patient, FE:

Examiner:	(*Pointing to her own reflection*) Who is this, next to the person [i.e. FE's reflection, which he takes to be a stranger in his house]?
FE:	I don't know.
Examiner:	Who does it look like? Have you seen this person in here before? (*pointing to the reflection of the examiner*).
FE:	That's you.
Examiner:	That's me?
FE:	Yes.
Examiner:	Me, here? (*pointing to herself*) What's my name?
FE:	I don't know, oh yes, it's Nora.
Examiner:	Nora, that's right. So that's me in the mirror?
FE:	Yes.
Examiner:	That's my reflection?
FE:	Yes.
Examiner:	And who is that? (*pointing to FE's reflection*).
FE:	I don't know what you would call him. It makes me a bit sick because he moves about freely with us. I don't be too friendly [sic] because I don't see it does him any good.[5]

What this exchange reveals is that despite the fact that the patient understands quite well what a mirror is and how it works (he is, after all, aware that he sees the examiner's reflection), he is simply unable to recognize *his own* reflection as such. But there is no reason to think that this is simply the result of a mistaken inference: if it were, his condition could be treated simply by getting him to think about things more carefully. Rather, the mirror sign disorder seems to reside at quite a different level, one which talk about "concepts" and "inferences" cannot really get a proper hold on.

2.3 the challenge

From the perspective of the Received View, the data presented in the two preceding sections can seem puzzling. And if they stood on their own, they might be merely that: a puzzle. But it is important to see the deep challenge they raise to RV when they stand together.

[5] Breen et al 2000, pp. 84-85.

One might, for instance, try to brush aside the concerns raised in §2.1 and insist that inferential capacities can indeed have these sorts of lapses, especially among very young children. Similarly, one might argue that the deficits of mirror sign patients are inferential after all, and simply insist that their tendencies toward mistaken inference are especially deeply-seated. But such moves ought to be the exception rather than the rule: when capacities for inference are treated as the sorts of things that can go on vacation for weeks at a time even as one's more covert behavior suggests that the answers have been reached, *and* that can go awry in a way that cannot be overriden by explicit knowledge, the sense that what one is talking about are *inferential* capacities begins to fade.

Alternatively, a defender of RV might be convinced by our interpretation of the "mirror sign", and so abandon the claim that RV provides a viable account of *mature* MSR, holding instead only that inferences are critically involved in the processes by which this capacity is initially *acquired*. This sort of *ad hoc* move might be motivated by the insistence that there is *some* sort of cognitive leap involved in learning this sort of self-recognitional capacity, and that an inference of some sort is the most likely candidate to fill this gap.

But it is important to recognize that the "most likely candidate" in the eyes of a theorist is not always, and perhaps not even usually, going to be the most ecologically efficient route. And so while we do allow that the capacity for MSR requires a good deal of cognitive sophistication, the view to be developed in the subsequent section treats it as a fundamentally *sensorimotor* capacity that is "scaffolded" on top of more basic perceptual skills. It is only with this alternative on the table that the difficulties we have raised for RV in this section can be fully appreciated.

3. An alternative

The task of this last section is to propose an alternative to the Received View, one which provides a robust explanation of successful MSR while also squaring with the studies canvassed in Section 2. According to our alternative proposal, MSR rests ultimately on sensorimotor capacities for keeping track of the way sensory stimulation varies with bodily movements. Crucially, the present discussion is meant to provide a "task-level"[6] description of what MSR involves; the specific details of its implementation will be left for further work. Here, we simply outline the proposal, briefly explain how it is meant to account for the data, and suggest a way to evaluate it.

[6] Or "personal (/animal) level". See McDowell 1994.

3.1 "information-based" behaviors

Thoughts and behaviors that are responsive to perceptual information can take one of two forms. On the one hand, they might rest on an *identification* of the perceived object (or event, or location, or …) as the object it is. So, for example, one might receive a phone call and, not recognizing who it is on the phone, ask for the person's name and conclude on that basis how to address him or her. Similarly, an animal might see a tree in the distance and employ a "cognitive map"[7] to conclude that it is the location of its home, and head in that direction for this reason. On the other hand, some thoughts and behaviors are directly *information-based*: the animal's grasp of the identity of the object to which the it is responding is fixed immediately by the animal's perceptual relation to the object.[8]

We can make the notion of information-based reference clearer by considering a more detailed example. Suppose an animal feels something pressing up against its back, and responds by running away, turning to look, playing dead, or whatever. Now, it seems *possible* that this feeling could have been an illusion. This does not entail, however, that in order to respond to this experience as it did, the animal need have reasoned:

1. I have a certain sort of feeling.
2. Usually this feeling is the result of something's pressing up against my back.
3. So, there is (probably) something behind me.

Indeed, if this sort of inferential progression were required, very few animals would make it in the wild. Additionally, the fact that the capacity to treat such sensory information appropriately is something that may need to be *acquired* does not entail that it is acquired through inference rather than, say, trial and error or even brute association. In other words, responding to sensory information that is in some sense ambiguous need not rest on deliberate disambiguation; it can instead be the product of a much more basic attunement to one's surroundings, and an *immediate* (and *non*-inferential) disposition to react appropriately to certain sorts of perceptual information.

[7] See Jacobs and Schenk 2003, Jacobs 2003.

[8] For the basic distinction between identification- and information-based thoughts, see Evans 1982, ch. 5. The extension of this treatment to the domain of *behavior* is, in our opinion, implicit in Evans's seminal work.

3.2 the proposal

The purpose of this philosophical interlude is hopefully clear enough: it is meant to suggest the possibility that an animal might come to be able to respond to the deliverances of mirrors in much the same way that it responds to more "ordinary" perceptual information, and that the process by which this ability is acquired might not be that different from the acquisition of more rudimentary sensorimotor capacities.

Consider, for example, that mirroring surfaces, as well as objects whose movements align with one's own, each have quite distinctive "looks", which vary in very specific ways as one's own body moves. With this in mind, the idea that there is a particular "sensorimotor profile"[9] associated with viewing oneself in a mirror seems compelling. But–and this is the key step–to the extent that this is so, it seems that an animal can recognize its mirrored reflection by being *immediately* sensitive to this sensorimotor profile: upon receiving perceptual information of a certain sort (namely, the sort that is usually involved when one views oneself in a mirror), the appropriate behavioral repertoire can be brought into play *directly*, thereby bypassing the putative need for a self-identifying inference of the sort described in §1.3.

Furthermore, the behavioral dispositions thereby invoked need not rest on any *explicit* awareness of the fact that the animal is looking at itself, any more than appropriate responsiveness to the feeling of something on one's back demands such reflective self-awareness (see §1.2). To say this is of course not to deny that such self-awareness is *sometimes* involved in MSR; it is only to point out that the relevant sort of "self-identification" can in principle be realized in one's behaviors even if it does not arise to the level of deliberate thought.

3.3 putting the embodied view to work

Crucially, the alternative view we have been developing does not deny that MSR is in some sense a "cognitively sophisticated" achievement. The key idea is only that such cognition can be realized in a way that is both *embodied*–insofar as it rests on immediate behavioral dispositions–and *embedded*–insofar as it involves a direct sensitivity to the structure of certain aspects of one's environment–rather than processes that are more narrowly "mental". As we will argue in this section, this "Embodied View" (EV) can account quite nicely for the findings we raised in Section 2 as challenges to RV.

Consider first the mirror sign. As Breen and her colleagues argue, there is reason to think that self-misidentifications of this sort rest on *perceptual* and *affective* disturbances: in the case of their patient FE, for example, the ability to

[9] For more on this notion of a "sensorimotor profile", see Noë 2004.

recognize his own face in a mirror was severely compromised, while the other patient they discuss was simply unable to recognize mirrored spaces as such.[10] This clearly fits well with EV's claim that it is a sensorimotor capacity that is at the basis of MSR.

Similarly, EV proposes that the acquisition of the capacity for MSR is fundamentally a matter of acquiring a battery of sophisticated sensorimotor dispositions. On this view, learning to self-recognize in mirrors will resemble somewhat the process of learning to hit a ball, and the piecemeal trajectory described by Courage and her colleagues indeed fits this prediction. A similar proposal has been put forward by the psychologist Katherine Loveland, on the basis of her work with autistic children:

> … the mirror is a kind of tool for mediated perception of things (including the self) that are not ordinarily visible from the observer's standpoint in the environment. Through a process of perceptual learning, the observer gradually learns what the mirror affords (seeing things not located where you are looking) and how to use one accurately (i.e., how to tell where the reflected thing is actually located and how to relate its location or movements to one's own).[11]

The notion of an "affordance" is especially helpful here because, at least on J.J. Gibson's classic view, the recognition of affordances does not rest on inferential leaps but is instead a matter of an immediate perceptual attunement to one's environment.[12]

In both of these areas, then, EV hopes to provide an account of MSR that is sensitive to those of its particularities that are difficult for RV to capture.[13] It remains to suggest a way in which further research might be able to determine which view should be preferred.

3.4 the next step

To test the respective strengths of EV and RV, we propose to employ the tools of *adaptive robotics*, in which virtual organisms are created and allowed to develop capacities to respond to their virtual environments with varying degrees

[10] Cf. also Ramachandran et al 1997 and Binkofski et al 1999, which indicate that "mirror agnosia" (the inability to recognize mirrored spaces) is the product of deficits in spatial representation.

[11] Loveland 1993, pp. 241-242.

[12] See Gibson 1979, ch. 8.

[13] Some of the more distinctively philosophical aspects of EV have been discussed in Schwenkler 2006.

of sophistication. Such virtual worlds provide *cognitive models* of the behaviors realized in them, which can then be used to develop and test specific predictions.[14]

In the present context, then, we propose to construct virtual worlds in which organisms whose cognitive profiles satisfy those in EV and RV respectively are given the opportunity to acquire the capacity for mirrored self-recognition. There are several crucial questions that such models will help us to answer: (1) Is it possible for organisms without reflective self-awareness or inferential abilities to acquire the capacity for MSR, as is suggested by EV? (2) If so, then what does the developmental trajectory of such a learning process look like, and does it more closely resemble the data described in §2.1 than the learning processes of virtual organisms on the model of RV? (3) If the capacity for MSR can be acquired along the lines of EV, then what is it about organisms that cannot acquire this capacity that accounts for their limitations?

Conclusion

The account of mirrored self-recognition provided by the Received View has considerable intuitive support, but further investigation reveals that it suffers from some deep problems. The Embodied View proposed here aims to account for MSR in a way that is at once more adequate to its real-life structure and more ecologically plausible. It is hoped that further investigation, both on actual subjects and via cognitive models in virtual environments, will confirm its predictions and enable it to be developed in more detail.[15]

References

Amsterdam, B. 1972. Mirror self-image reactions before age two. *Developmental Psychology* 5: 297-305.

Binkofski, F., G. Buccino, C. Dohle, R.J. Seitz, and H.-J. Freund. 1999. Mirror agnosia and mirror ataxia constitute different parietal lobe disorders. *Annals of Neurology* 46: 51-61.

Breen, N., D. Caine, M. Coltheart, J. Hendy, and C. Roberts. 2000. Towards an understanding of delusions of misidentification: Four case studies. *Mind and Language* 15: 74-110.

[14] See Stewart 2006 **[and Stewart, this volume??]**. For helpful software tools for computational cognitive modeling, as well as more in-depth discussions of the procedures involved, visit http://rob.ccmlab.ca:8082/ccm.

[15] Thanks are due to John Campbell, Brad Morris, Alva Noë, and Terry Stewart, as well as to audiences at Stanford University, University College London, and–of course – Cognitio 2006 for helpful comments and discussion.

Courage, M.L., S.C. Edison, and M.L. Howe. 2004. Variability in the early development of visual self-recognition. *Infant Behavior and Development* 27: 509-532.

Evans, G. 1982. *The Varieties of Reference*. New York: Oxford University Press.

Gallup, G.G. 1970. Chimpanzees: Self-recognition. *Science* 167: 86-87.

Gallup, G.G. 1977. Self-recognition in primates: A comparative approach to the bidirectional properties of consciousness. *American Psychologist* 32: 329-338.

Gallup, G.G. 1987. Self-awareness. In Mitchell, and Erwin (eds.), *Comparative primate biology, Vol. 2, part b: Behavior, cognition, and motivation* (New York: Alan R. Liss): 3-16.

Gibson, J.J. 1979. *The Ecological Approach to Visual Perception*. Boston: Houghton Mifflin.

Jacobs, L.F. 2003. The evolution of the cognitive map. *Brain, Behavior and Evolution* 62: 128-139.

Jacobs, L.F., and F. Schenk. 2003. Unpacking the cognitive map: The parallel map theory of hippocampal function. *Psychological Revew* 110: 285-315.

Loveland, K. Autism, affordances, and the self. In Neisser 1993: 237-253.

McDowell, J. 1994. The content of perceptual experience. *Philosophical Quarterly* 44: 190-205.

Mitchell, R.W. 1993a. Mental models of mirror-self-recognition: Two theories. *New Ideas in Psychology* 11: 295-325.

Mitchell, R.W. 1993b. Recognizing one's self in a mirror? A reply to Gallup and Povinelli, De Lannoy, Anderson, and Byrne. *New Ideas in Psychology* 11: 351-377.

Neisser, U. (ed.). 1993. *The Perceived Self: Ecological and Interpersonal Sources of Self-Knowledge*. New York: Cambridge University Press.

Noë, A. 2004. *Action in Perception*. Cambridge: The MIT Press.

Postal, K.S. The mirror sign delusional misidentification syndrome. In Feinberg and Keenan (eds.), *The Lost Self: Pathologies of the Brain and Identity* (New York: Oxford University Press, 2005): 131-146.

Ramachandran, V.S., E.L. Altschuler, and S. Hillyer. 1997. Mirror agnosia. *Proceedings of the Royal Society of London* B 264: 645-647.

Schwenkler, J.L. 2006. Mirrors, misidentification, and the sense of the self. Presented at the London-Berkeley Conference, University College, London.

Stewart, T. 2006. Dealing with complexity: The analysis of embodied behavior. Presented at Cognitio 2006, University of Quebec at Montreal.