The puzzle of mirror self-recognition

Johannes L. Brandl[1]

1 Introduction

Since the discovery that chimpanzees are able to pass a simple mirror test (Gallup 1970), much has been written about how self-awareness manifests itself in human and animal behaviour. When one removes an unusual mark on one's face that a mirror reveals, does the removal count as evidence for self-recognition? If it does, does it also count as evidence for the existence of a rudimentary self-concept and/or the capacity to ascribe properties to oneself? While Gallup has argued that a properly executed mirror test warrants a positive answer to these questions (Gallup 1982, 1998; Gallup et al. 2014), others have interpreted the mirror task as indicating only a sensorimotor capacity to match visual and kinaesthetic stimuli (see Mitchell 1993a; Heyes 1994; Povinelli and Cant 1995; Povinelli 2000; Schwenkler 2008). Further evidence from developmental studies suggests that success in the mirror task may be indicative of a socially induced form of self-awareness that emerges from integration of both perceptual and interpersonal experiences (see Rochat 2001; Rochat and Zahavi 2011). None of these interpretations has found general acceptance, however.

In this paper, I argue that in order to advance the debate on mirror self-recognition, two critical moves need to be made. First, we need to overcome the common sense intuition that recognizing oneself in the mirror is an all-or-nothing phenomenon. It is a well-established empirical fact that "self-recognition does not emerge suddenly with one particular behaviour, but develops gradually" (Bertenthal and Fischer 1978, p. 44). This observation holds not just for self-recognition in general, but also applies specifically to the phenomenon of *mirror* self-recognition. Like other cognitive achievements, recognizing oneself in a mirror requires some practice and therefore allows for

---

✉ Johannes L. Brandl
 Johannes.brandl@sbg.ac.at

[1] Department of Philosophy, University of Salzburg, Franziskanergasse 1, A-5020 Salzburg, Austria

Springer

"partial accomplishments" (see Courage et al. 2004). Second, we must allow that the process of learning to recognize oneself in a mirror may vary from species to species. While experiments with non-human animals may demonstrate the proficiency to track how sensory stimuli vary with bodily movements, similar experiments with children may reveal something else about what goes on in the mind of a 2-year-old child when it responds meaningfully to its mirror image. Accordingly, the apparently contradictory views mentioned above may turn out to be compatible to some extent.

Once mirror self-recognition is conceived of as a gradual phenomenon, there is conceptual space for a notion of *partial* self-recognition that precedes the ability to fully understand the nature of one's mirror image. But what does it mean to partially recognize oneself in a mirror? How can we make sense of this *prima facie* puzzling idea? Perner's seminal work (Perner 1991) suggests that an important skill children acquire during their second year is the ability to use multiple models in representing facts that do not fit into their reality model. Numerous studies confirm that this ability is a necessary prerequisite for children's pretend play, synchronic imitation, and early empathic responses (see Asendorpf and Baudonnière 1993; Bischof-Köhler 1988, 2012). Given the fact that children manifest these abilities at the same age when they first pass a mirror test, their behaviour in front of mirrors may express their capacity to operate with multiple models.

The goal of this paper is to explain children's early success in the mirror task with a multiple-models theory of partial mirror self-recognition. One of the principal advantages of this explanation is that it allows us to integrate otherwise incompatible views about the mirror task. Accordingly, the first part of the paper reviews the controversy about Gallup's interpretation of the mirror task and explores what lessons one can learn from that debate. Section 1 sets the stage by offering a piecemeal definition of mirror self-recognition. Section 2 takes a critical look at a rich, mentalistic interpretation of the mirror test, while Section 3 explores the limits of a more parsimonious interpretation based on the sensory-kinaesthetic matching hypothesis. Section 4 engages with the scepticism about the mirror task that prevails in the philosophical literature and rebuts what I call "The Master Argument against the mirror test." Part two of the paper, starting with Section 5, introduces the multiple-models hypothesis for explaining why mirror self-recognition is developmentally related to pretend play. Section 6 elaborates the hypothesis and answers objections against taking a capacity for pretence to play a role in the mirror task. Finally, Section 7 shows how the concept of partial self-recognition also sheds light on the affective qualities of mirror experiences when children are stunned or embarrassed to see their faces.

2 Setting the stage

What happens cognitively when we recognize ourselves in a mirror directly facing us? A funny story told by the physicist and philosopher Ernst Mach reveals the complexity inherent in this well-known phenomenon (see Mach 1886). Once Mach stepped on a bus and thought he saw a shabby schoolmaster doing the same, only to recognize a few moments later that he was looking at himself in a mirror. If we conceive of this event in slow motion, we can see it unfold in four steps: (1) seeing someone that looks like a shabby schoolmaster, (2) taking this person to step on the bus at the very same moment,

(3) recognizing the awkwardness of the coincidence and suspecting that there may be no other person there, and finally (4) recognizing that one sees one's own mirror image.

Analysing the phenomenon further, we may define the concept of mirror self-recognition in a piecemeal way, as follows:

A subject S recognizes herself in the mirror if and only if S comes to know in some way or other

(i)   that the perceptual experience is not to be taken at face value,
(ii) that the source of the visual impression of another person is oneself,
(iii) that the appearance of another person provides information about oneself,
(iv) that one is looking at an image of oneself.

The definition leaves open *how* the subject comes to know what the four clauses specify. Some of this knowledge may be purely procedural knowledge, while some of it may be declarative knowledge and require the formation of explicit beliefs about oneself. Regardless, as long as a subject does not explicitly form the belief that she is seeing herself in the mirror, we do not have a case of mirror self-recognition in the full sense. Conceiving of mirror self-recognition in this way resolves what Rochat refers to as the "self-other paradox", which requires "the suspension of the normal social experience of others facing you with eye contact. Mirror reflection of the self is paradoxical in the sense that what is seen in the mirror is the self as another person" (Rochat 2001, p. 205).

To notice and resolve the self-other paradox, one must step back and reflect on the awkward nature of the described experience. This raises the question as to whether there may be a pre-reflective form of self-recognition that avoids the paradoxical situation of identifying oneself with what looks to be different person. With the help of our piecemeal definition, we can see how this might be possible. Each clause in the above definition describes a new insight that adds something to what the subject has already discovered. Thus, it may be that an animal or a child realizes that there is something awkward about the reflective images in mirrors without realizing that it can control these images. Gaining such control may therefore count as a pre-reflective or naive form of self-recognition. Such naivety is no longer feasible once a child begins to wonder about her control over the mirror image. Indeed, she may doubt her own competence and feel anxious about observing a scene while also being part of it. It may then take several steps to overcome this confusion until a child fully grasps how she can stand in front of the mirror and at the same time be the person in front of her that is looking at herself.

It thus becomes clear that the phenomenon of mirror self-recognition can be tested at different levels. Gallup's famous mirror mark test targets this complex phenomenon at a more basic level than Mach's story. What this test reveals may be only the naive form of self-recognition that subjects possess when they satisfy conditions (i) and (ii). Or it may reveal something about the confusion that arises when children become wary of their own mirror image, or the trajectory that enables children to reach the level of full self-recognition. Since these possibilities are often not spelled out, one easily gets the impression that "the research literature [on the mirror test] is full of contradictions" (Bertenthal and Fischer 1978, p. 44). But these apparently contradictory views may simply turn out to be accounts of different aspects of what is a highly complex phenomenon.

In taking this liberal stance, I try to avoid prejudging things by what would be a merely terminological decision to use the term "self-recognition" in either this way or that. We can thereby avoid asking the wrong questions; there is no point in asking questions that eventually turn out to be unanswerable. Thus, instead of asking:

Q1: Can one demonstrate self-recognition with the mirror test?

a better question is:

Q1*: Which components of full self-recognition can be tested with the mirror test?

One also cannot expect an answer to the following question:

Q2: Does mirror-self recognition imply the presence of a self-concept?

We should rephrase the question in this way:

Q2*: Which components of full mirror self-recognition involve the use of a self-concept?

Admittedly, simpler formulations are often more attractive. It may seem tedious to translate sloppy uses of the term "self-recognition" into a more precise idiom, as it is tedious to replace straightforward questions like Q1 or Q2 by more cumbersome questions like Q1* and Q2*. A more precise idiom, however, enables us to determine what we can learn from seemingly incompatible interpretations of the mirror task.

3 The mentalistic interpretation

In his extensive writings on the mirror test, Gallup advocates an interpretation of the test according to which it reliably indicates whether an agent recognizes itself in the full sense (Gallup 1970, 1977, 1982, 1994, 1998).[1] For Gallup, passing the test means to detect an unusual change in one's appearance. This requires a self-concept and awareness of the fact that one's appearance changes over time. Since such awareness involves an active use of memory about oneself, it also means that one has at least some insight into one's own mental capacities and one's existence in time. This leads Gallup to regard the mirror test as a proof that self-awareness, consciousness, and mind are not "mutually exclusive cognitive categories" (see Gallup 1982, p. 237).

Such far-reaching conclusions require strong empirical support. Gallup therefore makes much of the fact that a mirror test, as he conceives it, must follow a rigorous procedure. By changing the procedure, Gallup fears that one makes it too easy for animals to pass the test. For example, from the fact that elephants, dolphins, or even pigeons successfully perform *some* test with mirrors, it does not follow that they

---

[1] But Gallup comes close to acknowledging that the mirror test may target only a weaker sense of self-recognition; he says that chimpanzees (but not monkeys) "realized that their behaviour was the source of the behaviour being depicted in the reflection" (Gallup 1994, p. 35).

possess a self-concept.[2] Only chimpanzees and orangutans have passed a rigorously performed mirror test, Gallup holds. What, then, is the procedure on which Gallup bases this claim?

Gallup's study with chimpanzees consisted of observational and testing phases. In the first phase, he observed that some chimpanzees showed a peculiar interest in mirrors and used it to inspect body parts that otherwise they could not see. Taking this to be a sign of self-recognition, Gallup hypothesized that the chimpanzees would notice some clearly visible change in their appearance, like a rouge mark on their face. After allowing the chimps to remain acquainted with mirrors for 10 days, he anaesthetised them and marked their faces with an odourless dye above one eyebrow and on one ear. When the animals fully recovered, Gallup reintroduced mirrors into their cages. He then recorded all movements that seemed to indicate the animal noticing marks on its face. Relevant movements included touching the marked areas, examining them closely in the mirror, and trying to remove the markings. Gallup also controlled for mark-directed movements when the animals could not see their mirror image. If there was a significant increase in mark-directed behaviour in front of the mirror, the animal passed the test.

Gallup's experimental approach supersedes all prior findings about animals' and humans' behaviour in front of mirrors. Yet it is far from clear that the testing procedure with all its control conditions warrants the conclusion that there is "a qualitative psychological difference among primates," namely a capacity for self-recognition that "may not extend below man and the great apes" (Gallup 1970, p. 87). There are many questions that need to be addressed before one can accept that conclusion. Why, for example, is it important that animals are first familiarized with mirrors before taking the test? What happens cognitively during this initial phase of the experiment? Do they form memory images of what they look like, which later enables them to discover unusual marks as diverging from their normal appearance? Or do they need this training period merely to get used to the presence of mirrors and suppress aggression stimulated by them? Another concern is their level of motivation. It could well be that animals recognize changes in their appearance yet feel no compulsion to respond to them; they might simply lack any such motivation. Furthermore, there may be alternative explanations for animals' failure to respond to mirror marks in the way Gallup expects. Finally, one might raise a question about the different motor skills employed by elephants, dolphins, birds, and so on. Can one dismiss these differences when conducting mirror tests, or does success depend on specific forms of bodily self-awareness that may differ markedly in different species?

The lesson here seems to be that not all of Gallup's specific measures may be essential for properly executing a mirror test, while independent evidence is needed to support the psychological conclusions he draws from the data. This becomes even more vivid when one performs a mirror test with young children. While the procedure is basically the same, researchers usually skip the initiating phase because most children

---

[2] Animal studies employing some version of a mirror test have produced many surprising results that are not easy to explain. It is not clear, for example, why not all Great Apes seem to be able to pass such tests. Gorillas often do not respond to their mirror image even after long exposures, whereas corvids can be rather easily trained to do so. (See Parker et al. 1994; Keenan et al. 2003, and Gallup et al. 2011.)

are already accustomed to mirrors in their environment even if they have not yet learned to see themselves in mirrors. Few reports exist regarding how children or adults respond when confronted with a mirror for the very first time. (I will discuss one such case in Section 7.) Additionally, researchers do not anesthetize children, but mark their faces while they are awake in a way designed to disguise the manoeuvre. Children may nonetheless notice the markings at a subconscious level, and this may influence the results collected during the testing phase. Furthermore, there are notable exceptions to the claim that children normally pass the test between the age of 18 and 20 months (to which I will return below).

Nevertheless, there are psychologists who agree with Gallup's heuristic, even if they do not accept his specific conclusions about the minds of animals. They agree, in other words, that "it may be heuristically valuable to suppose that organisms capable of recognizing themselves in mirrors are creatures that can conceive of themselves" (Gallup 1998, p. 240). We see this heuristics at work in Lewis and Brooks-Gunn's comment about a 15-month-old boy: "Self-recognition implies to us that David has a notion of self, a notion that he has existence in a particular point in space and that this existence has characteristics or features" (Lewis and Brooks-Gunn 1979, p. 6).[3]

Thus we are left with a dilemma. If one follows Gallup's presumption that self-recognition involves possession of a self-concept, then it becomes questionable whether the mirror test really shows that self-recognition *in this sense* takes place. To avoid such scepticism, one may revert to a less demanding notion of self-recognition that does not require all the controls that Gallup specifies. But then one must admit that this more basic form of self-recognition may be found much more widely in the animal kingdom and much earlier in human development than has been assumed.

Various attempts have been made in the literature to distinguish between conceptual and non-conceptual, reflective and pre-reflective, or explicit and implicit forms of self-awareness (see e.g. Bermùdez 1998). While such distinctions may be useful in a broader investigation of self-awareness, I do not think that a distinction between explicit and implicit forms of self-recognition could resolve the dilemma at hand.[4] First, these distinctions may not be equivalent and they all rest on highly controversial presuppositions. Second, as we have already seen, there is a more direct way how we can distinguish between different levels of mirror self-recognition; we do not need such general distinctions for introducing a concept of partial mirror self-recognition. And thirdly, how one understands the notion of explicit self-recognition largely depends on whether one takes it as implying possession of a self-concept. The case of chimpanzees might be a reason to rethink such an implication. That is to say, we might stick with the view that nonhuman primates do not possess a self-concept, but grant chimpanzees more than an implicit self-awareness if it turns out that they realize more about the nature of the mirror image than a purely behavioural interpretation of the mirror test allows. In order to assess this possibility, however, we must first consider what a purely behavioural interpretation amounts to.

---

[3] For a more recent defence of this claim, see Savanah 2012. For a critical reply, see Mitchell 2015.
[4] For similar concerns about employing such distinctions, see Musholt 2015.

4 The behavioural interpretation

In this section, I examine an alternative interpretation of the mirror test that tries to explain success with more basic cognitive abilities than those involved in possessing a self-concept (see Heyes 1994, 1995, 1996; Mitchell 1993a, b, 1997, 2003; Povinelli and Cant 1995, and Povinelli 1995, 2000; Schwenkler 2008). The conclusions these authors draw from the mirror task can be interpreted in different ways. On a weaker reading, their view provides a critique of Gallup's mentalistic interpretation, but leaves the reliability of the mirror test unchallenged. On a stronger reading, the critique amounts to a denial that the mirror test is a test for self-awareness or for self-recognition at all.

According to Mitchell, success in the mirror test is enabled by a kinaesthetic-visual matching mechanism that is either missing or disturbed in animals that fail the test (Mitchell 1993a, b, 1997). Povinelli & Prince take a similar line when they explain an interesting dissociation between young children's competence in responding to their mirror image and incompetence in recognizing themselves on videos:

> "[O]ur explanation begins by denying that self-recognition has anything to do with understanding the reflective property of mirrors… [T]he typical development of a domain-general system of representation in 18–24 month old children provides them a means of forming relations between objects and events in the world and schemes held in mind. […] [T]he ability to form such equivalence relations allows the organism to understand that the representation of self held in mind (an immediate, on-line representation of the self's actions, movements, and possibly desires) is equivalent to the image they see in the mirror. […] This does not require the organism to understand that the mirror is an accurate and contingent image of itself." (Povinelli and Prince 1998, p. 52 f.)

In further studies, Povinelli demonstrated how limited an apparent capacity for self-recognition can be by introducing a temporal delay (Povinelli 2001).[5] He filmed 18–24 month-old children while their faces were secretly marked. After a few minutes, he showed them videos of the marking event and then observed how they responded to these videos. Although most of these children passed a standard mirror test, hardly any of the children could use the information on the videos to detect the mark on their own face. If children are unable to process the information presented in video recordings, this raises some serious questions about their live performance. Do these children possess a capacity for self-recognition that they can exercise only in certain conditions, or do they merely manifest a behavioural skill that enables them to behave as if they were recognizing themselves in a mirror? Since the latter possibility cannot be ruled out, success in a mirror test does not warrant an inference to self-recognition, since

---

[5] More recent studies by Suddendorf confirm Povinelli's data and provide further evidence about children's difficulties with recognizing themselves in videos. (See Suddendorf 1999.)

they still lack the domain general system of representation required for recognizing oneself in a mirror.[6]

Heyes takes this scepticism further when she attacks the methodological legitimacy of Gallup's procedure. She declares any evidence provided by the mark test to be "no more compelling than anecdotal reports of mirror self-recognition," and contends that research in this area has "led to a considerable waste of time and resources," since the test is not a "valid measure of mirror self-recognition" (Heyes 1995, p. 1541, Heyes 1996, p. 470). In contrast to those who think that videos may provide us with a clearer view of when self-recognition sets in (see Eddy et al. 1996), Heyes doubts that videotaping instead of directly observing test subjects would bring us any closer to a reliable test for self-recognition (see Heyes 1995, p. 1538). In this connection, she mentions the following task in which not only chimpanzees, but rhesus monkeys are quite proficient: using a joystick to track moving objects on a screen with a cursor. To pass this test, the animals do not need to know anything about the technology that links their movements of the joystick with the movements of the cursor. All they know is how "to use the cursor position as a source of novel, displaced, visual feedback on the position of its hand" (Heyes 1994, p. 917). Heyes contends that in a similar way, animals might learn to use the information they receive from a mirror image. Contrary to what Gallup concludes about the inability of monkeys to pass the mirror test, then, Heyes predicts that sooner or later, "evidence of mirror-guided body inspection will be found in both chimpanzees and monkeys" (ibid.)

Thus, Heyes recommends that one should stop using the term "self-recognition" in connection with the mirror test, and instead refer to it as a form of mirror-guided behaviour. Notice, however, that this conclusion is warranted only if one takes "mirror self-recognition" to mean self-recognition in the full sense. Heyes tacitly makes this assumption when she presumes that self-recognition requires a second-order representation of oneself:

> "[I]f it were shown, using a properly designed experiment, that an animal can use a mirror to derive information about its body, we would not automatically have evidence that the animal 'recognizes' itself, or even its body, in the mirror. […] to use a mirror in this way it would not be necessary for the animal to recognize, or form a second-order representation of, the representational relationship between the mirror and its body. In principle, the animal could simply use the information available in the mirror image, without having any thoughts about why the information is available or useful." (Heyes 1996, p. 472.)

What Heyes describes here is another form of the dilemma mentioned earlier: Either we take the term "self-recognition" to mean self-recognition in the full sense, in which case we can attribute such a cognitive achievement to an animal only if it possesses the capacity for second-order representations. Or we interpret the capacity to pass the mirror test as a kind of practical skill, in which case there is no good reason why we

---

[6] Povinelli and Cant 1995 speculate that a domain-specific skill of matching stimuli might explain the extraordinary climbing technique of orangutans called "clambering". In the same vein, one might speculate that a domain specific skill in using cameras might enhance children's comprehension of their own video recordings.

should describe this skill as a form of self-recognition at all. But why accept this dilemma in the first place? Why not allow for the possibility that some animals recognize themselves in a mirror in a "naive" way without making use of second-order representations?[7]

It is not within my expertise to evaluate the empirical plausibility of Heyes' predictions. But it requires no such expertise to see that her radical scepticism about the mirror test goes too far. We should not forget that the same task may pose different demands for animals and for human children. If you ask a layman and an expert to describe an archeologic site, for example, the former may find the task quite easy because he does not recognize subtleties about which the archeologic expert is acutely aware. Something similar could be true for the mirror task. Young children who pass a mirror test may have to cope with the illusory experience of seeing another person in the mirror. This could make the task considerably more difficult for them than for animals that, by having become accustomed to mirrors, no longer experience this illusion. This consideration shifts the burden of proof onto the sceptic; for the sceptic must explain why monkeys fail a mirror test that should be no more difficult for them than the cursor-tracking task that Heyes describes. Even if these tasks may be similar, engaging with one's mirror image may involve an extra difficulty that specific species might simply overlook or ignore, while others cope with it directly.

5 A master argument against the mirror test

Let me now turn to the philosophical debate about mirror self-recognition, in which a certain form of scepticism has become the prevailing attitude. Drawing on the empirical considerations just reviewed, the sceptical position is that the mirror test tells us very little about the nature of human self-awareness.[8] The sceptic's reasoning takes different forms depending on what background assumptions are in play. However, there seems to be convergence on a general argument that I would like to make explicit. I call it the Master Argument against the mirror self-recognition test:

1. Self-recognition requires an explicit form of self-awareness.
2. Visual-kinaesthetic matching does not imply an explicit form of self-awareness.
3. Passing the mirror test requires only visual-kinaesthetic matching.
   Therefore: Passing the mirror test is not a reliable indicator of self-recognition.

As one can see, the notion of "explicit self-awareness" figures prominently in this argument. That is also it's weakness, I will argue, because the force of the argument rests on an equivocation as to what explicit self-awareness means. Taken in one sense, the conclusion of the argument becomes incontestable, while taken in another sense, the argument can easily be blocked by rejecting its third premise. One can deny this premise because, as mentioned before, passing the mirror test can mean different things

---

[7] The suggestion to take an embodied approach to mirror self-recognition points in the same direction. (See Schwenkler 2008.)

[8] In contrast to this attitude among philosophers, Gallup takes the critique lightly and sees no reason to change his view, (see Gallup and Povinelli 1993; Gallup 1994; Gallup et al. 2011, 2014).

for subjects with different capacities. Subjects with an explicit self-awareness may not be able to pass the test without noticing that they are in full control of the mirror image they produce, while subjects that lack such an explicit self-awareness may pass the test merely by using a sensory matching mechanism.

Let me illustrate this general observation with discussions of mirror self-recognition by Lynn Rudder Baker, David Papineau, and Tyler Burge. All of them take a sceptical stance towards the mirror test. If my criticism of the Master Argument is sound, this consensus needs to be reconsidered.

Baker discusses mirror self-recognition as a potential problem for her distinction between what she calls "weak" and "strong" first-person phenomena (Baker 2000, p. 60). In the first category she puts the subjectivity of experience, distinguishing it from the ability to represent oneself as oneself—a strong first-person phenomenon. The problem for Baker arises from the observation that "chimpanzees' self-recognition seems to fall between weak and strong first-person phenomena" (ibid. p. 62). To maintain a sharp line between subjective phenomena and genuine cases of self-consciousness, she contends that it is a necessary condition for any form of self-consciousness that one be able to represent oneself as oneself (ibid. p. 67 and p. 69). This leads her to take the following position on the mirror task: although chimpanzees can recognize themselves in a mirror, this does not provide any evidence for self-consciousness or possession of a self-concept. In response to this, it may be said that Baker's entire argument depends on a very strong interpretation of 'explicit self-awareness'. Why should one accept on such an interpretation the first premise of her argument?

In a similar vein, Peacocke takes issue with Gallup's test as demonstrating possession of a full-blown self-concept. The same points that weaken Baker's criticism therefore might also apply to Peacocke's reasoning. Yet, in contrast to Baker, Peacocke acknowledges a form of self-awareness that does not require a self-concept. With that possibility in mind, Peacocke offers a more nuanced account of the mirror task:

> "It is not at all obvious that the kind of self-consciousness manifested by passing the mirror test is restricted to creatures who possess concepts. […] Intuitively, the kind of self-consciousness required to pass the mirror test involves taking a third-person perspective on oneself, while appreciating that it is oneself who is present in that third person perspective. That is something that could in principle be done at the nonconceptual level." (Peacocke 2014, p.192)[9]

In order to get clear about what Peacocke proposes here, we would need to know how the distinction between conceptual and non-conceptual self-consciousness maps onto the four components in my piecemeal definition of self-recognition. Presumably, a form of self-recognition that does not involve all four components would count as a form of non-conceptual self-consciousness. But then the Master Argument would not go through, because such non-conceptual self-awareness requires more than a capacity for matching visual and kinaesthetic stimuli. That is to say, by saving premise (1) of the argument, Peacocke pays the price of making premise (3) false.

---

[9] While such a non-conceptual self-awareness may be involved, it is not necessarily required, as Peacocke points out when he says that it may "very reasonably be objected that the ability to use a mirror in removing a mark from one's forehead can be explained without attributing to the animal recognition of itself in the mirror" (Peacocke 2014, p. 191).

Finally, consider what Tyler Burge has to say about the visual-kinaesthetic matching mechanism that enables animals to pass a mirror test:

"…connections among kinaesthetic proprioception, the proprioceptive body schema, and visual perception [allow] the perceiving individual to associate ego-centric indexes with the same individual as *perceived from the perspective of different perceptual modalities*. The perceived individual is perceived as having ego-significance, in this limited way, from different perspectives." (Burge 2013, p. 151 f.)

Tellingly, Burge mentions in this context neither that the mirror test is supposed to be a test for self-recognition nor that self-recognition is supposed to imply a form of explicit self-awareness. The omission of these background assumptions may suggest different things. It might suggest that for Burge the mirror test does not test for self-recognition at all, as the Master Argument says. However, Burge might also want to suggest that self-recognition requires a level of self-awareness that transcends the kind of information processing that underlies perceptual experience. I am inclined to follow this latter interpretation. The perceptual skills that Burge mentions are clearly not sufficient for an explicit kind of self-awareness. Hence, at least in the human case, where we know that such forms of explicit self-awareness exist, we cannot assume that children pass a mirror test without recognizing themselves in some way. That means, once again, that at least in the human case premise (3) of the Master Argument fails.

As these examples from the philosophical literature illustrate, the Master Argument only appears to be a strong argument against the mirror test. In fact, the conclusion of the argument is either obvious, or the argument can be resisted by questioning its third premise, depending on what we presuppose – or know from independent evidence – about the self-awareness of an agent taking a mirror test. Now that this argument is out of the way, I will try to advance the debate by proposing an interpretation of the mirror test that avoids the shortcomings of both mentalistic and behaviouristic interpretations.

6 The multiple-models theory of partial mirror self-recognition

While Gallup tested chimpanzees for self-recognition, it was found that children pass the same test beginning at the age of 15–18 months (Amsterdam 1972). In studies with children it is usually taken for granted that "passing the test" means to progress from seeing another person to recognizing oneself in the mirror. Many developmental researchers assume that children are capable of making this large step by the age of 18 months: "Full self-recognition of a mirror image as their own occurs unambiguously in most infants at about 18 months of age when they respond to a spot of rouge that has been covertly applied to their nose" (Courage et al. 2004, p. 510). Nevertheless, there are notable exceptions to the claim that children normally pass the test between the age of 18 and 20 months. In cultures where children are highly submissive to adults, they hesitate much longer or do not try to remove the mark from their face (see Broesch et al. 2010). Relatedly, children tend to behave differently when they notice that other people have been marked in the same way (see Rochat et al. 2012). These results remind us that it seems to be an incredibly large step from (erroneous) other-recognition to a full blown self-recognition. Are 2-year-old children really capable of taking this step?

In answering this question, it will be crucial to take a closer look at other cognitive capacities that develop at nearly the same time. More specifically, it is significant that children who pass mirror tests also show the first signs of empathic responses (Bischof-Köhler 1988, 1991), pretence and invisible displacement (Perner 1991; Gergely 1994), synchronic imitation (Asendorpf and Baudonnière 1993; Asendorpf 2002), and personal pronoun use (Lewis and Ramsay 2004).[10] These correlations are hardly accidental—and they motivate taking a more general approach in how we explain children's success in the mirror task.

According to Perner's influential theory, a common factor that explains why these achievements evolve together is the use of multiple models (Perner 1991). When children incorporate all of the information that they receive from their environment into a single model, their conception of reality remains limited to what they can presently perceive and what is consistent with their present perceptions. It is only when children learn to construct alternative models that their reasoning takes them beyond what is presently given. Multiple models enable them to compare their present situation with memories of past events so that they can construct models of hypothetical situations. They also enable children to learn to play different roles by imagining themselves in another person's situation. These are the primary achievements attained during the second year of life which indicate that a major development in children's cognitive abilities has occurred.

Early cases of mirror self-recognition in infants may turn out to be yet another manifestation of the ability of using multiple models. In what follows, I will take this idea one step further and argue that in order to make sense of the partial character of self-recognition in 2-year old children, we may presume that children also invoke their newly acquired capacity for engaging in acts of pretence when they are puzzled by seeing their own mirror image. That said, I will first briefly consider how the multiple-models theory works in the case of pretence, invisible displacement, and empathy.

If an adult pretends to be a dog, a 1-year-old child observing his behaviour might not know what to make of it. According to her model of reality, this behaviour is unusual and incomprehensible. In order to make sense of the pretence, the child must construct an alternative model in which barking noises do not only sound like the barking of a dog, but also count as such. Now the child can approach the person playing a dog and engage with him (or rather with the person playing the dog), using what she knows about real dogs. She can act in an as-if mode, without devoting any thought to why the other person acts as she does or what she is trying to achieve by it. Once the child switches to an alternative model, the person playing a dog becomes a dog in the play-world, and the child has to switch back to her reality model to make sure that the person playing the dog is still a human person.

Consider next invisible displacement. What needs to be explained here is why 2-year-old children, but not 1-year-olds, search for an object that is placed under a cover while children observe the hiding. Children restricted to the use of a single model fail this test, Perner suggests, because they continuously update their model of reality and

---

[10] How the formation of a self-concept ties in with the use of the first person pronoun is a highly problematic question that must be left open here. On this issue, even foremost experts do not seem to have a settled view. For instance, Rochat contends that "it is feasible that self-concept emerges under the pressure of growing linguistic competence," but then corrects himself by saying that "although language and explicit self-concept appear connected in the timing of their emergence, it does not mean that they are mutually dependent." (Rochat 2001, pp. 192–193).

ignore information that is inconsistent with it. What they see from this restrict-ed perspective is that an object is first present and then no longer present. From this information alone they cannot infer the invisible location of the object. They begin to comprehend what is going on only when they use two models to represent the sequence of events. While continuously updating their model of reality, children also construct an alternative model with information retrieved from memory—a model that is inconsistent with their current model of reality. In this way, they can represent the object that is presently invisible as still existing in the alternative model. The contrast between these two situations (object visible / object not visible), may prompt them to start looking for the object close to where they last saw it.

In a similar vein, Perner's theory also explains why children do not show signs of empathy before they learn to use multiple models. Children are born with a tendency to suffer from emotional contagion. It is for this reason that they get upset and start crying when they hear other babies cry or notice their distress. As long as they cannot resist such contagious emotions, children are not in a position to empathize with others by offering help or comfort. They are themselves in need of being consoled. Perner contends that children need multiple models to overcome this problem. If a child perceives that another child's teddy bear is torn open, she can filter out this information by locating it in a mental file that concerns only the other child and not herself. Distancing herself from the stressful experience in this way relieves the child from getting upset; and this enables her to develop empathic responses (see Perner 1991, pp. 132 ff.).

All this shows that multiple models are a powerful tool that may explain a wide variety of achievements, including pretend play, searching for invisible objects, and empathy. How then could it also explain children's success in the mirror task? In taking this step, it is crucial to draw a distinction between self-recognition in the full sense and what is merely a partial form of self-recognition. The idea is not completely new, but is usually not spelled out in sufficient detail to see how it works. Courage presents the idea only briefly in this short remark:

> "In the case of mirror self-recognition, for example, the child may be coming to terms with two "selves" that are normally viewed as one – the already present and familiar self perceived through the temporal contingency of visual-proprioceptive information and the one they see reflected in the mirror who 'looks and moves like me'. (Courage et al. 2004, p. 529)

The crucial question that Courage only touches here concerns the "partial accom-plishment,", as Haith and Benson (1998) call it, that lies in between not recognizing oneself at all and being fully aware of the fact that one sees a reflection of oneself in the mirror. I suggest that we may describe this phenomenon as follows:

(i)   The child is still in the grip of seeing another child.
(ii) It does not believe, but pretends that there is another child in front of her.
(iii) It realizes that by using her normal skills it can control the duplicate body in the mirror.
(iv) It is motivated to act such that the other child looks normal, or is good-looking, charming, etc.

Each of these conditions is crucial for explaining the special significance of the mirror test. Condition (i) explains why the situation in front of the mirror may be more difficult to comprehend for a child than for a chimpanzee that is no longer seeing a conspecific in the mirror. As in the case of perceptual illusions, children may not be able to completely suppress the impression that some other person is there, perhaps staring back at them. Not being able to fully comprehend the situation either, they might come up with the "trick" of using multiple models for cognitively resolving the illusory experience. The trick consists in pretending, instead of believing, that one sees another person in the mirror. This is what condition (ii) suggests. Condition (iii) describes a skill that animals may also possess when they figure out how to gain control over the image that they no longer perceive as the image of a conspecific. Finally, condition (iv) is crucial for explaining why children are motivated to respond to the discovery of the secretly applied mark if they do not identify themselves with the person wearing the mark. When they first touch the mark (on their own face, not in the mirror), this may be explained as a way of testing whether they are still in control of their duplicate body. Hence they know that they did not mark themselves. Removing the mark may then be explained by a desire for their duplicate body to look normal. Children also sometimes pose in front of the mirror, behaviour that may also be explained by a desire for their duplicate body to look strong and attractive. I will return to the affective dimension of a mirror experience in Section 8, after some further clarifications of my proposal.

7 Further clarifications

So far I have argued that Perner's multiple-models theory leads to a plausible explanation of early mirror self-recognition; and I have connected this with the idea that children at this age might pretend, but not believe, to see another person in the mirror. Now it is time to address some worries that this proposal raises. In particular, I need to say more about the conceptual requirements that an act of partial self-recognition involves. Does it require possession of a minimal self-concept? If so, what might such a concept amount to? Relatedly, I will address the worry that pretending to see another person in the mirror is a much more complex task than what the multiple-models theory suggests. It also may be objected that since there are some obvious differences between acts of pretence and an act of mirror self-recognition, those differences speak against my account of partial self-recognition. Finally, I will address the objection that the account I am offering is artificial and that there may be more natural ways to implement a theory of partial self-recognition.[11]

Does partial self-recognition require a minimal self-concept? This question arises because it seems necessary, on the present proposal, for a child to have such a concept in order to respond by touching herself after discovering a mark on her face via a mirror. In order to explain this fact, the multiple-models theory seems to require that a child has to represent herself as existing in different models. That is how Perner

---

[11] I am happy to take up these objections here that reviewers have raised against an earlier version of this paper. It also has been suggested that my proposal is ad hoc and lacking empirical support. This objection I take to be unjustified given what I have said about the need for a notion of partial self-recognition and how it fits in with the data that support a multiple-models theory.

describes from a first person perspective the discovery that one would make in taking a mirror test:

> "In order to understand that the person in the mirror is myself, I have to be able to construct two models, one in which I figure as myself in reality and another in which I figure as myself in the situation reflected in the mirror." (Perner 1991, pp. 132f.)

If children took this reflective stance, the following sequence of thoughts would have to occur in a child's mind:

1. In the situation reflected in the mirror, I have a mark on my face. (Model 1)
2. The situation reflected in the mirror is the same as the situation in reality.
3. I have a mark on my face in reality. (Model 2)

Those who think that 2-year-old children already possess a minimal self-concept might find nothing objectionable in ascribing such thoughts to them. Such a response would be in line with interpretations of the mirror task requiring a form of "self-objectivation" (see e.g. Bischof-Köhler 1991). It would also accord with the view that mirror-recognition shows that children conceive of themselves as objects in public space (see Rochat and Zahavi 2011). However, none of these claims follows from the multiple-models theory alone.

These claims do not follow because the multiple-models theory does not require children to take the reflective stance that Perner describes. Perner is far from endorsing Gallup's rich, mentalistic interpretation. He agrees with Heyes and Povinelli that one can pass a mirror test without understanding the mirror as a representational medium (see ibid., p. 132, and p. 133fn.). Thus it is not clear that one needs a self-concept to update one's reality-model after seeing oneself in the mirror. Nor is it clear that one needs to understand that the situation reflected in the mirror and the situation depicted in one's reality-model are identical. These are insights that may come only much later when one achieves full self-recognition.

This leaves us with the question as to how we might articulate the content of the two models that the child constructs without using the first-person idiom. How can we explain the relation between the two models if the child does not represent herself as a part of both models? If a child does not yet possess a self-concept, obviously it is difficult to make explicit the thoughts that might cross her mind during a mirror task. One might despair of this problem and revert to the idea of "immediate self-recognition" that requires no inferential process at all (see Schwenkler 2008). But I think we can do better than that.[12]

The child may use one model for representing what people usually look like and another model for representing the unusual person that the child sees in the mirror. As in a pretend play, information put into the second model is not used to update the child's reality model. As such, the child does not see herself, but rather a funny looking

---

[12] Schwenkler may be right, however, that pathological failures to recognize oneself in a mirror may best be explained as a deficit arising on the sensorimotor level and not from a failure to draw inferences of a certain kind.

person in her mirror-world. If the child wants to bring the mirror-world in correspondence with reality, i.e. with what people normally look like, it may act in accordance with the following inference:

1*.   This person has something unusual on her face.
2*.   This person should be like other people.
3*.   This person will look normal if my face is cleaned.

Since the first-person idiom is still needed in the final step of the child's reasoning, one might still suspect that a self-concept must be operative if the child cleans her own face. At this stage, however, the use of the pronoun "my" in step (3*) does not carry the same significance as the pronoun "I" in an act of self-reflection. Specifically, it does not signify that the child has formed a self-image that contrasts with her mirror image and thereby indicates that there is something wrong with her appearance. Such self-reflection presupposes that the child identifies itself with the person it sees in the mirror; and we have yet no reason to assume that such an identification takes place.

I consider it to be a major advantage of the multiple-models theory that it can show us how to avoid ascribing explicit first-person thoughts to those who pass a simple mirror test. The kind of first-personal reasoning that is required for relating the two models may be explained at a procedural level without ascribing even a minimal self-concept to a 2-year-old child. Nevertheless, the cognitive skills of children passing the mirror test cannot be reduced to the simple coordination of perceptual stimuli. While such coordination may explain an animal's success in the mirror task, it may turn out that not only the possession of a self-concept, but also the use of multiple models is a distinctively human achievement. Nothing is prejudged about a crucial difference between animals and humans here, however. If it turns out that such an intermediary approach is also useful in the study of mirror self-recognition in animals (see Swartz 1997), we might have to reconsider whether some animals have the skill to build a model around their mirror image and use it for engaging in an act of pretence. This would not threaten, but strengthen my proposal, because it would give the multiple-models hypothesis a further domain of application.

How can one pretend to see another person in the mirror? If an adult person sees herself in the mirror, she may not like what she looks like and therefore imagine seeing someone else. Such cases of self-deception may take different forms, depending on whether it is just a playful thought or whether the person really comes to believe that what she sees is not herself. In either case, it may seem far-fetched to claim that something similar happens when a 2-year-old child only partially recognizes herself in a mirror.

Much depends at this point on how one conceives of the act of pretence that occurs in a 2-year-old child when it passes a mirror test. It is not the case, on my view, that such pretence begins with the reflective first-personal question: "Is this really me?" In addition, I do not want to ascribe to children a form of self-deception that involves self-alienation when they pretend that there is someone else in the mirror. Pretending to see some other person is not the same as pretending not to see oneself. The former act is much simpler and leads to merely partial mirror self-recognition. That said, it need not be as simple as acting like a dog. One may compare it with a form of "acting as if" that occurs when children pretend that a block of wood is a car. In this case children do not behave as if the block

were a real car; otherwise they would have to do a number of irrational things, like trying to get inside the block. In the case of dogs, a cognitive capacity to switch between multiple models seems to suffice, whereas the latter case seems to require a minimal (if nascent) understanding of the appearance/reality distinction.

At this point, it may seem that my account of partial mirror self-recognition will fail for either of two reasons. The first is that children do not actually recognize themselves in the mirror, as the multiple-models theory describes; they only behave as if this were the case. This would align the present proposal with similar scepticism about self-recognition in animals; but I do not think that the parallel holds. In the case of animals, such scepticism is reasonable because *all* of the evidence that we have about animal behaviour seems compatible with the hypothesis that they do not even partially recognize themselves in mirrors. This is not so in the case of children, as we have seen. Indeed, there is converging evidence that children at the age of two begin to become aware of their own agency and relationship to others. Consequently, it would require significant argumentation to make plausible the claim that children are capable of no self-recognition at all. Nothing that I have claimed suggests that the multiple-models theory could be used to support such scepticism. That a child does not fully identify herself with the person in the mirror does not mean that she behaves only as if she does.

The second objection is that children who can pretend that a block of wood is a car might already be capable of self-recognition in the full sense. Leslie has argued that such acts of pretence require a form of meta-representation and an innate module for a "theory of mind" (see Leslie 1994). Yet this is far from clear, as I have argued elsewhere (Brandl 2012). Children can learn to use multiple models before they learn to reason flexibly about how different models can represent what is real and what is not. Thus I see no reason why children should not be able to pretend to see another person in the mirror well before they understand the representational properties of mirrors, hence also before they are able to recognize themselves in a mirror in the full sense.

Let me dwell on this critical point a little further. What evidence is there that pretence in the mirror task may occur without a self-concept and without the capacity for meta-representation? Although I cannot offer direct evidence for this claim, there is compelling indirect evidence. If children could not engage in such pretence, we could not explain how a sensory-kinaesthetic matching mechanism might enable children to pass the mirror test. Consider, for instance, the following description of sensory-kinaesthetic matching by Gergely:

> "We can assume that the [child] placed in front of the mirror has a primary model of reality that contains representations of external objects as well as a representation of the self. […] When confronted with a mirror, with experience the [child] learns to map the visual information in the mirror onto its primary model of reality. […] However, this perfect correspondence relation breaks down when the visual features of the image of the nonvisible parts of the body are mapped onto the representation of the corresponding object in space. […] The systematic correspondence between the primary model of reality and the mirror image can be salvaged only by attributing to the self-representation the mismatching visual features of the corresponding object in the mirror." (Gergely 1994, p. 56)

Gergely freely speaks of a "representation of the self" and of "self-representation" to make clear that a child has to realize that something unusual happened *to itself*. The child is supposed to realize this indirectly by noticing a mismatch between its normal self-representation and the visual features of the corresponding object in the mirror. When she notices this mismatch, the child attributes to itself the unusual feature (the mark) by adding it to its self-representation; and this is supposed to explain why it reaches up to touch or remove the mark that it has discovered in the mirror.

But is it necessary for the child's primary model to contain a representation of itself (if not of its self)? Consider the parallel that Gergely draws between success in mirror self-recognition and success in the task of invisible displacement, when children use a secondary representation (such as a memory image of the object disappearing) to update their primary representation of reality. In both cases, Gergely suggests, a child is "using information from sources other than direct perception" (Gergely 1994, p. 55). No further conceptual differences need to be invoked to distinguish invisible displacement and mirror self-recognition from a simple visual-kinaesthetic matching mechanism. This means that from a conceptual point of view, passing a mirror test with the help of multiple models may not be more demanding than passing it by using a sensory-kinaesthetic matching mechanism. These are, of course, different mechanisms; but neither of them presupposes a self-concept or the capacity for meta-representation.

Differences between pretend play and mirror self-recognition. While there may be a common faculty that children use in both pretend play and passing a mirror test, it may still seem unlikely that children actually pretend to see another person in the mirror. There seem to be clear disanalogies between mirror self-recognition and pretend play that cast doubt on this claim. First, in mirror self-recognition children do not express the same kind of fun that they express in pretend play. Second, nothing in pretend play corresponds to the child's motivation to remove the mark from her own face. Third, why do children show no surprise when they are able to remove unusual marks from their own face?

Although these differences are undeniable, I do not think that they speak against my account. Using multiple models is a very basic capacity that opens up a wide range of possibilities for getting beyond reality. Some of these possibilities may be pure fun for the child; some may be frightening; and cases of pretence may involve a mixture of both. Mirror experience seems to be a mixed case because it involves both voluntary and involuntary components. The presence of the mirror image is something with which the child has to cope in one way or another. She is not free to make it up, as she is free to make up the fantasy that there is a dog barking or that a wooden block is a car. Additionally, she also has a limited capacity for making the mirror image disappear, for example by changing her position or turning around, because the image will reappear as soon as the child faces the mirror. It may therefore take some courage to interact with this image as long as one does not fully understand its nature. This may explain why children lose the playfulness that they first exhibit when trying to engage with the "stranger in the mirror". To overcome anxieties associated with the uncanny stranger, it is helpful for children to discover that they can change the appearance of the mirror image by drawing grimaces or make it move in accordance with their own movements. The fact that it takes significant effort to turn the presence of the mirror image into a pretended object may explain why there are no signs of fun or surprise left, once the child has gained control over it's mirror image.

Is an act of pretence really necessary for partial self-recognition? Finally, one might raise the objection that while the idea of partial self-recognition seems valuable, one does not need a story about pretence to implement this idea. Why not use the multiple-models theory without adding this problematic feature? If mirror self-recognition works like invisible displacement, pretence does not seem to play an essential role. Some proposals in the literature on mirror self-recognition point in this direction, but they all seem to be missing something, which suggests that an appeal to pretence is not superfluous in this context.

Jens Asendorpf discovered an interesting correlation between children's ability to engage in synchronic imitation and their ability to pass the mirror test (see Asendorpf and Baudonnière 1993; Asendorpf et. al. 1996; Asendorpf 2002). Like Perner, he suggests that in both cases children form a primary representation of reality and then use a secondary representation to keep track of their own activity. This enables them to detect whether another person is imitating their own behaviour, and then imitate the other person in turn to maintain the synchrony. Asendorpf describes such imitation as a form of communication in which children "enjoy the reciprocity inherent in their joint play" (Asendorpf 2002, p. 63). He goes on to offer the following explanation of mirror self-recognition:

> "Mirror self-recognition requires coordinating a mirror-image (primary representation) with one's representation of oneself (the "Me"; James 1890). This latter representation is a secondary representation because it is not a perceptual reality. Instead it is a constructed mental image of oneself that can be manipulated in fantasy." (ibid., p. 66)

Like others, Asendorpf assumes that children can pass the mirror test only when they represent their own activities as part of an explicit self-representation. In contrast to Gergely, however, Asendorpf does not take this self-representation to be a primary representation, but considers it to be a "constructed self-image", i.e. a secondary representation that is not part of the child's perceived reality.

What this shows, I think, is that merely distinguishing between different models – or representations, as Gergely and Asendorpf would put it – does not suffice to explain how these models can be used in so many different tasks. Crucially, children also need to have some understanding of how these different models are connected. At a minimum, they must keep track of what remains constant from model to model, something that a simple mechanism of switching from one model to the other does not guarantee. That there always are some constant features follows from the fact that the construction of a model always depends on what children know about reality, by either modifying their reality model or enriching it with additional information. Sometimes this enrichment will accord with what they know about reality; for instance, when they reach for an object that is currently not visible. But sometimes constructing a model requires generating conflict with what the reality model contains. This is so in the case of both mirror self-recognition and pretence, if one pretends a situation that does not obtain. A child sees a flat, shiny surface (model 1) which displays a figure that looks like a person standing behind the surface. She then transforms this perception into an image of what becomes her pretended duplicate body (model 2). The skill required

to do this is quite sophisticated, because it requires combining the reality of the mirror with the non-real character of the person it presents. I cannot see how a child could solve this problem – before fully coming to understand the nature of the mirror image – without engaging in an act of pretence.

Another effort to propose a "moderate explanation" of visual self-recognition has been made by Thomas Suddendorf and David Butler. They also invoke a multiple-models theory to steer a middle course between Gallup's rich interpretation and the lean accounts offered by its rivals (Suddendorf and Butler 2013, 2014). They present the core idea of their proposal as follows:

> "[T]he test reflects the capacity to generate and compare multiple models of the same thing. By comparing an expectation about one's physical appearance with current perceptions of a reflection, inconsistencies, such as the mark, can be noted and motivate exploration." (Suddendorf and Butler 2013, p. 122)[13]

Here Suddendorf and Butler provide a plausible answer to the question as to why children should be motivated to remove a mark from their forehead. To wit, they must have an expectation of what they normally look like and share a motivation to restore "normality", when they discover the unusual mark in the mirror. If children are indeed surprised, or even shocked by the mark, this may be the right explanation. But as I mentioned before, shock or surprise are not necessary features of early mirror self-recognition. For this reason, we also need an explanation why children remove the mark from their forehead when we have no reason to ascribe to them an expectation of what they normally look like. On my account, what suffices to explain this behaviour is a general expectation what *other* people normally look like and a preference for the child's duplicate body to look like a normal person. In this way, we need not assume that children have the constant expectation of not wearing funny marks on their face. Instead, we explain why they remove the mark from their own head by the fact that they lack any motivation to pretend that somebody is wearing such a mark.

Pointing out these differences in explaining mirror self-recognition may seem tedious; but solving the puzzle of self-recognition hangs precisely on these differences. There is no reason to attribute to children self-representational abilities that would allow them to perceive themselves in a mirror like adults do. My contention is that children at the age of two are not yet able to do so. They achieve something quite different which is even more remarkable: they keep attending to the mirror while seeing in it a child in whose existence they no longer believe. Without fully comprehending the situation, they use a trick of pretence to get a grasp of and thereby control over the initially strange figure. This achievement is so extraordinary that it deserves its own description: *partial self-recognition*.

---

[13] One must be careful here not to confuse the "collating of representation", as Suddendorf & Butler call it, with a visual-kinaesthetic matching mechanism that Mitchell postulates in his explanation of the mirror task. The former concept requires children to use multiple models, and is therefore cognitively more demanding than a simple matching mechanism that may be found in many species. Musholt seems to overlook this difference (see Musholt 2015, p. 150).

8 The affective dimension of the mirror experience

Since the invention of the mirror test, the question has been discussed whether emotional responses may be indicators as to when children begin to evaluate and hence recognize themselves in a mirror. The emotional responses that seem relevant range from excitement or playfulness to coyness, frustration, aggression, and embarrassment, but also include forms of pleasure and pride (see Amsterdam 1972; Lewis and Brooks-Gunn 1979; Rochat 2003, 2009; Lewis 2011). In this final section, I want to show how the present theory can shed light on some of the emotional responses that young children exhibit when they become acquainted with their mirror image.

Rochat and Zahavi claim that this is the most serious problem for any theory of self-recognition that focusses on representational skills (see Rochat and Zahavi 2011). How can we explain why the mirror experiences has its "uncanny" nature and why children avoid seeing themselves in the mirror or "freeze" in an embarrassed state in front of their mirror image? I will not be able to address these questions in full detail. The following remarks should suffice to show, however, that cognitive explanations like the one offered in this paper can meet this challenge to some extent. First, I will attempt to dispel the worry that the theory I have proposed underestimates young children's performance in mirror tasks by disregarding their emotions. This complaint would be justified only if children's emotions were *helpful* in recognizing themselves in a mirror. But this does not seem to be the case. Second, I will show how the multiple-models theory may actually provide a plausible explanation for the confusion that children seem to feel when they are puzzled by their own mirror image.

There are three kinds of affective responses elicited by the mirror test: responses typical for children that do not yet pass the mirror test; responses typical for children that fully understand the nature of the mirror image; and responses typical for children that are in a transitory phase between not recognizing themselves and fully recognizing themselves.[14] In the first category we find signs of excitement, wariness, and aggression. These responses may be explained by the fact that children responding in this way have unrealistic expectations about what the mirror image indicates to them. They might want to engage with a playmate, get excited about doing so, and then become insecure or aggressive when their expectations are frustrated. The third category could include all affective responses that children might share with adults when they observe themselves in a mirror, including feelings of pride and shame. There is reason to assume, however, that at the age of two such complex emotions can exist only in a very rudimentary form. We may therefore expect the affective responses of children passing the mirror test during their second year to be found mostly in the second category. Two-year-olds might experience the tension of not knowing how to respond or feel embarrassed about the effect that their own movements have on the person in the mirror. They may also feel relief or joy after resolving this tension by utilizing multiple models and a trick of pretence.

---

[14] This is a very rough distinction that would certainly need to be refined in a fuller treatment of this topic. It is worth mentioning, though, that this is tripartition does *not* coincide with another tripartition suggested by Bischof-Köhler. She distinguishes between children not passing the test (non-recognizers), children passing the test (recognizers) and children that she classifies as "transitional" (see Bischof-Köhler 1988). While for Bischof-Köhler most two-year-old children are clear recognizers, they are still in the second category according to my scheme, because they do not yet fully comprehend the nature of the mirror image.

What I have not yet mentioned are basic kinds of "social emotions" that according to Rochat and Zahavi play a central role in mirror self-recognition, even before children are capable of experiencing pride and shame. Such social emotions generate the existential feeling of being "exposed" to the view of others and of being alienated:

> When I am seeing myself in the mirror, I am seeing myself as others see me. I am also seeing myself as if I was another, i.e., I am adopting an alienating perspective on myself. The enigmatic and uncanny character of the specular image is precisely due to this intermingling of self and other." (Rochat and Zahavi 2011, p. 209)

Rochat and Zahavi rely here on the work of phenomenologists like Merleau-Ponty, whose view on this matter they summarize as follows:

> "Merleau-Ponty's central idea is that mirror self-recognition exemplifies a troubled form of self-knowledge. To recognize oneself in the mirror does not simply involve an identification of the felt me which is here, and the perceived me which is there, it also and more importantly involves the dawning realization that the felt me has an exterior dimension that can be witnessed by others […]. In short, the decisive and unsettling impact of mirror self-recognition is not that I succeed in identifying the mirror image as myself. Rather, what is at stake here is the realization that I exist in an intersubjective space. I am exposed and visible to others." (ibid. p. 209)

To which of the three aforementioned categories does the experience that Merleau-Ponty describes here belong? Rochat and Zahavi contend that feelings of exposure and alienation are manifest as soon as children pass a mirror test, if not before. But in making this claim, they may overestimate the emotional capacities of children at this early age.

One example discussed by Rochat and Zahavi does not seem to support their view. This is the exceptional case of a Russian boy that was deliberately kept away by his parents from all mirror surfaces until he was 2 years old. When Svyto finally was given the opportunity to see himself in the mirror, he embarked on an emotional "roller coaster", as Rochat and Zahavi describe it on the basis of the film that documents the case:

> "After a period of puzzlement and hesitation, Svyto first tries to engage his own image in playful exchanges, getting very excited in fencing and banging the mirror, exploring then the perfect contingency of his own movements and the movements of the specular image experienced. He eventually shows great distress and despair in not being able to create a dialogic exchange with the image. […] In a last sequence, after approximately 40 min, the movie shows Svyto coming to terms with his own reflection. […] relishing his own specular image, admiring and combing his hair with obvious delight." (ibid., p. 210, fn. 6.)

This description mentions no feeling of exposure or alienation. Generalizing from this case, one can raise two crucial questions about the role of emotions in children's

mirror experience: (1) How far might unpleasant emotions hinder children from recognizing themselves in a mirror? (2) How far might positive feelings help them eventually see themselves in the mirror image? The view advocated in this paper suggests the following account.

According to the multiple-models theory, a rather subtle change in children's emotional experiences is to be expected when they change their attitude towards the person they see in the mirror. They no longer *believe* that they see another child in the mirror, but go on to pretend that it is there. In this way, they lose the cognitive basis for emotions related to this belief, while retaining the opportunity for an emotional engagement with a person they do not take to be real.

Precisely how this pretence modulates their emotions is of course difficult to say. Whatever those emotions might be like, however, they will hardly resemble the feelings of a person that takes it to be a matter of fact that she is looking into her own eyes when she stares into a mirror. This once again highlights the fact that children who pass the mirror test at the age of two are still in a transitory phase. They have moved beyond the stage when they are simply disturbed by the stranger in the mirror, but they are still a step away from understanding why there is no one there. In this transitory state, the emotions that children experience may be confused emotions.[15] They are certainly confused in their verbal responses, as the following report by Povinelli about his 3-year old daughter testifies. After seeing herself in the mirror, Jennifer expressed her puzzlement thus: "It's Jennifer, but why is she wearing my shirt?" (Povinelli 1995, p. 81).

The report about Svyto also makes sense from this perspective. It shows a boy in the transitional phase between not recognizing and recognizing himself in the mirror; and this is why his emotions are so varied and volatile. Even when Svyto finally seems to learn that he is seeing himself in the mirror, it is not clear that he actually admires *himself*. Playing a game in which he engages with a strange duplicate would also explain his joy. In coming to terms with his mirror image, he might enjoy seeing a happy boy in the mirror; and this may help him calm down and become happy. Such positive emotions, in turn, may help the boy eventually recognize himself as himself. In this way, an account of partial self-recognition offers a plausible account of the emotional confusion characteristic of children's early mirror experiences.

Have we then finally solved the puzzle of mirror self-recognition? Given the exploratory remarks in this final section, this claim would certainly be too strong. When it comes to answering questions Q1* and Q2* from Section 1, however, it seems that a solution is clearly in sight. The answer is just not as straightforward as the simpler questions Q1 and Q2 would allow. Yet this is just as it should be. We should not expect to reach a solution that dispels the magic of mirror experience. Indeed, we still want to indulge in the mystery when we deliberately pretend *not* to see ourselves in the mirror or be the shabby person we see. Without acknowledging the peculiarity of such experiences, the puzzle of mirror self-recognition will remain unsolved.

---

[15] Lewis distinguishes between what he calls "exposure and evaluative embarrasment" (see Lewis 2011, p. 125). Whether these categories appropriately capture the kind of embarressment experienced by a young child that is embarrassed about seeing herself in the mirror, must remain open at this point.

References

Amsterdam, B. (1972). Mirror self-image reactions before age two. *Developmental Psychobiology, 5*, 297–305.

Asendorpf, J. B. (2002). Self-awareness, other-awareness, and secondary representation. In A. N. Meltzoff & W. Prinz (Eds.), *The imitative mind* (pp. 63–73). Cambridge: Cambridge University Press.

Asendorpf, J. B., & Baudonnière, P. M. (1993). Self-awareness and other-awareness: mirror self-recognition and synchronic imitation among unfamiliar peers. *Developmental Psychology, 29*, 88–95.

Asendorpf, J. B., Warkentin, V., & Baudonnière, P. M. (1996). Self-awareness and other-awareness II: mirror self-recognition, social contingency awareness, and synchronic imitation. *Developmental Psychology, 32*, 313–321.

Baker, L. R. (2000). *Persons and bodies: A constitution view*. Cambridge: Cambridge University Press.

Bermùdez, J. (1998). *The paradox of self-consciousness*. Cambridge: MIT Press.

Bertenthal, B. I., & Fischer, K. W. (1978). Development of self-recognition in infant. *Developmental Psychogy, 14*, 44–50.

Bischof-Köhler, D. (1988). Über den Zusammenhang von Empathie und der Fähigkeit, sich im Spiegel zu erkennen. *Schweizerische Zeitschrift für Psychologie, 47*, 147–159.

Bischof-Köhler, D. (1991). The development of empathy in infants. In M. E. Lamb & H. Keller (Eds.), *Infant development: Perspectives from German speaking countries* (pp. 245–273). Hillsdale: Lawrence Erlbaum.

Bischof-Köhler, D. (2012). Empathy and self-recognition in phylogenetic and ontogenetic perspective. *Emotion Review, 4*, 40–48.

Brandl, J.L. (2012). Pretend play in early childhood: The road between mentalism. In M. Beran, J. Brandl, J. Perner, & J. Proust (Eds.) *The foundations of metacognition* (pp. 146–166). Oxford University Press.

Broesch, T., Callaghan, T., Henrich, J., Murphy, C., & Rochat, P. (2010). Cultural variations in children's mirror self-recognition. *Journal of Cross Cultural Psychology*, 1–13.

Burge, T. (2013). *Cognition through understanding*. Oxford: Oxford University Press.

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

Eddy, T. J., Gallup, G. G., & Povinelli, D. J. (1996). Age differences in the ability of chimpanzees to distinguish mirror-images of self from video-images of others. *Journal of Comparative Psychology, 110*, 38–44.

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. (1982). Self-awareness and the emergence of mind in primates. *American Journal of Primatology, 2*, 237–248.

Gallup, G. G. (1994). Self-recognition: Research strategies and experimental design. In S. T. M. Parker, W. Robert, & M. L. Boccia (Eds.), *Self-awareness in animals and humans* (pp. 35–50). Cambridge: Cambridge University Press.

Gallup, G. G. (1998). Self-awareness and the evolution of social intelligence. *Behavioral Processes, 42*, 239–247.

Gallup, G. G., & Povinelli, D. J. (1993). Mirror, mirror on the wall which is the most heuristic theory of them all? A response to Mitchell. *New Ideas in Psychology, 11*, 327–335.

Gallup, G. G., Anderson, J. R., & Platek, S.M. (2011). Self-recognition. In S. Gallagher (Ed.) *The Oxford handbook of the self*. Oup Oxford.

Gallup, G. G., Platek, S. M., & Spaulding, K. L. (2014). The nature of visual self-recognition revisited. *Trends in Cognitive Sciences, 18*, 57–58.

Gergely, G. (1994). From self-recognition to theory of mind. In S. T. M. Parker, W. Robert, & M. L. Boccia (Eds.), *Self-awareness in animals and humans* (pp. 51–60). Cambridge: Cambridge University Press.

Haith, M.M., & Benson, J.B. (1998). Infant cognition. In *Handbook of child psychology.* (vol. 2, pp. 199–254). New York: Wiley.

Heyes, C. M. (1994). Reflections on self-recognition in primates. *Animal Behaviour, 47*, 909–919.

Heyes, C. M. (1995). Self-recognition in primates: further reflections create a hall of mirrors. *Animal Behavior, 50*, 1533–1542.

Heyes, C. M. (1996). Self-Recognition in primates: irreverence, irrelevance and irony. *Animal Behavior, 51*, 470–473.

Keenan, J., Gallup, G. G., & Falk, D. (2003). *The face in the mirror: The search for the origin of consciousness*. New York: Harper Collins.

Leslie, A. M. (1994). Pretending and believing: issues in the theory of ToMM. *Cognition, 50*, 211–238.

Lewis, M. (2011). The origins and uses of self-awareness or the mental representation of me. *Consciousness and Cognition, 20*, 120–129.

Lewis, M., & Brooks-Gunn, J. (1979). *Social cognition and the acquisition of self*. New York: Plenum Press.

Lewis, M., & Ramsay, D. (2004). Development of self-recognition, pronoun use, and pretend play during the 2nd year. *Child Development, 75*, 1821–1831.

Mach, E. (1886). *Beiträge zur Analyse der Empfindungen*. Jena: Fischer.

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.

Mitchell, R. W. (1997). A comparison of the self-awareness and kinesthetic-visual matching theories of self-recognition: Autistic children and others. In J. G. Snodgrass & R. L. Thompson (Eds.), *The self across psychology. Self-recognition, self-awareness, and the self-concept* (pp. 39–62). New York: Annals of the New York Academy of Siences.

Mitchell, R. W. (2003). Subjectivity and self-recognition in animals. In M. R. Leary & J. P. Tanguey (Eds.), *Handbook of self and identity* (pp. 567–593). New York: Guilford Press.

Mitchell, R. W. (2015). A critique of Stephane Savanah's "mirror self-recognition and symbol-mindedness. *Biology and Philosophy, 30*, 137–144.

Musholt, K. (2015). *Thinking about oneself*. MIT Press.

Parker, S. T. M., Robert, W., & Boccia, M. L. (1994). *Self-awareness in animals and humans*. Cambridge: Cambridge University Press.

Peacocke, C. (2014). *The mirror of the world: Subjects, consciousness, and self-consciousness*. Oxford: Oxford University Press.

Perner, J. (1991). *Understanding the representational mind*. Cambridge: The MIT Press.

Povinelli, D. J. (1995). The unduplicated self. In P. Rochat (Ed.), *The self in infancy: Theory and research*. Amsterdam: Elsevier.

Povinelli, D. J. (2000). *Folk physics for apes. The chimpanzee's theory of how the world works*. Oxford: Oxford University Press.

Povinelli, D. J. (2001). The self: Elevated in consciousness and extended in time. In K. Skene & C. Moore (Eds.), *The development of the extended self in preschool children: Theory and research* (pp. 73–94). Cambridge: Cambridge University Press.

Povinelli, D. J., & Cant, J. G. H. (1995). Arboreal clambering and the evolution of self-conception. *The Quarterly Review of Biology, 70*, 393–421.

Povinelli, D. J., & Prince, C. G. (1998). When self met other. In M. Ferrari & R. R. Sternberg (Eds.), *Self-awareness. It's nature and development* (pp. 37–107). New York: The Guilford Press.

Rochat, P. (2001). Origin of self-concept. In G. Bremner & A. Foge (Eds.), *Blackwell handbook of infant development* (pp. 191–212). Oxford: Blackwell.

Rochat, P. (2003). Five levels of self-awareness as they unfold early in life. *Consciousness and Cognition, 12*, 717–731.

Rochat, P. (2009). *Others in mind. Social origins of self-consciousness*. Cambridge: Cambridge University Press.

Rochat, P., & Zahavi, D. (2011). The uncanny mirror: a re-framing of mirror self-experience. *Consciousness and Cognition, 20*, 204–213.

Rochat, P., Broesch, T., & Jayne, K. (2012). Social awareness and early self-recognition. *Consciousness and Cognition, 21*, 1491–1497.

Savanah, S. (2012). Mirror self-recognition and symbol-mindedness. In *Biology and philosophy* (pp. 657–673).

Schwenkler, J.L. (2008). Mental vs. embodied models of mirrored self-recognition: Some preliminary considerations. In B. Hardy-Valeé, & N. Payette (Eds.), *Beyond the brain: Embodied, situated, and distributed cognition*. Cambridge Scholars Press.

Suddendorf, T. (1999). Children's understanding of the relation between delayed video representation and current reality: a test for self-awareness? *Journal of Experimental Child Psychology, 72*, 157–176.

Suddendorf, T., & Butler, D. L. (2013). The nature of visual self-recognition. *Trends in Cognitive Sciences, 17*, 121–127.

Suddendorf, T., & Butler, D. L. (2014). Response to Gallup et al.: are rich interpretations of visual self-recognition a bit too rich? *Trends in Cognitive Sciences, 18*, 58–59.

Swartz, K. (1997). What is mirror self-recognition in nonhuman primates, and what is it not? In J. G. Snodgrass & R. L. Thompson (Eds.), *The self across psychology. Self-recognition, self-awareness, and the self-concept* (pp. 65–82). New York: Annals of the New York Academy of Siences.