

Mirrors, Misidentification, and the Sense of the Self

1. This paper is about what it is to conceive of oneself as a person. In particular, I'm going to focus on what we can call *self-reference*: the ability to refer to oneself in thought, and *self-identification*: the sort of process that makes it so that, in any given case, such thoughtful reference latches onto oneself rather than something else.¹ The very specific kinds of self-reference and self-identification I'm immediately concerned with are those that involve our use of *mirrors* to monitor how things are with ourselves; I'm going to discuss what is involved in this sort of 'mirrored self-recognition' and the thoughts² that ensue from it, and argue that certain traditional (and admittedly rather intuitive) models of how this proceeds can't account for some of the central phenomena that attend to it. What is perhaps the crucial moral of this paper can be summed up as follows: our capacity for skillful mirrored self-recognition, and the important place mirrors occupy in our everyday self-monitoring, shows that even in what can seem to be one of the most Cartesian of activities – referring to ourselves in thought – we humans are often *embodied and embedded*. What this means will become clearer as I go along.

2. Let's begin, though, by examining an amusing scenario. In the Marx Brothers movie *Duck Soup*, Harpo tricks Groucho by disguising himself to look like him, and carefully mimicking his movements as if he were his mirror image. (I don't recall if Groucho ever quite realizes that something is amiss.) Now suppose for a moment that Harpo has a hair out of place, but Groucho does not. Seeing what appeared to be his image in a mirror, we can imagine that Groucho would think to himself, 'I have a hair out of place', or something of the like, and perhaps try to smooth it out. Clearly this is an instance of a false belief: Groucho has no hair out of place. But *someone* does, and Groucho has that person plainly in view: so if, for whatever reason, Groucho had concluded that '*Someone or other* has a hair out of place', or something of the like, this belief would have been true.

With regard to this latter possibility, two questions arise: first, would this existentially general belief have been expressive of *knowledge* that someone had a hair out of place?; and second, *why* would Groucho have decided to stick with this open-ended conclusion rather than settling on himself as the subject? On a bit of reflection, it turns out that these questions are obviously related: for whether a belief amounts to knowledge is very closely tied to the way it is arrived at. So the story needs to be filled out a bit more.

Well, we could imagine that Groucho had caught on to Harpo's trick, and so formed this existentially general belief out of suspicion. In this circumstance, it seems intuitive that his belief does indeed amount to knowledge: he knows that the disheveled person is either himself (if his suspicions are misplaced and it's a mirror after all) or someone else (whoever is trying to trick him). But suppose this *wasn't* the reason why Groucho formed the existentially general belief: perhaps instead it was formed simply on the basis of his belief that *he* had a hair out of place ('*Fa*, so $\exists xFx$ '). In this latter case, it seems unlikely that even this new belief, while true, would have amounted to knowledge: after all, he formed it on the basis of a falsehood, and he thought all along that the 'someone' in question was himself.³

One other point should be noted: if Harpo does a good enough job of mimicking Groucho, then the *only grounds* Groucho will have to conclude that 'Someone is thus' based on what he sees of Harpo (and takes to be himself) will proceed by way of his grounds for concluding 'I am thus'. This is because Groucho, like any mature human being, understands – though perhaps only implicitly – that when one stands in front of (what seems to be) a mirror and sees (what seems to be) a reflection, then – so long as the appearance of the reflection is not very different from one's own appearance, and so long as its movements keep in tune with one's own – it is *oneself* that one sees (in the foreground, anyway), and not someone else. Hence seeing (what seems to be) one's mirrored reflection does not, unless the possibility of various Harpo-like scenarios is brought into play, seem to give one any immediate grounds to form beliefs about the appearance of anyone other than oneself.

In short, the moral I want to draw from this story is that when one observes (what seems to be) one's appearance in (what seems to be) a mirror, one's claim to know any open-ended facts about the appearance of 'someone or other' on this basis will generally stand and fall with one's claim to know the specifically first-personal facts about *one's own* appearance. In other words, the beliefs one forms about oneself based on seeing (what seems to be) one's reflection in (what seems to be) a mirror are what philosophers have called *immune to error through misidentification* with respect to the first-person singular pronouns: it is not possible that what one sees fails to give one sufficient grounds for knowing that 'I am thus' while still giving one sufficient grounds for knowing that 'Someone or other is thus'.⁴

3. This may be a startling conclusion, so it will be worth examining it a bit further before drawing out its consequences. Let's start with the passage in the *Blue Book* in which Wittgenstein inaugurates the recent discussion of immunity to error through misidentification. He begins by distinguishing 'two uses of the word "I" (or "my")', which he calls respectively 'the use as object' and 'the use as subject'. As examples of the first category of uses he suggests 'My arm is broken' and 'The wind blows my hair about'; as examples of the latter he gives us 'I see so-and-so' and 'I have a toothache'. He then proposes that we can mark the difference between these categories as follows:

The cases of the first category involve the recognition of a particular person, and there is in these cases the possibility of an error, or as I should rather put it: The possibility of an error has been allowed for. ... It is possible that, say in an accident, I should feel a pain in my arm, see a broken arm at my side, and think it is mine, when really it is my neighbor's. ... On the other hand, there is no question of recognizing a person when I say I have a toothache. To ask 'are you sure it's *you* who have pains?' would be nonsensical.⁵

The basic point should be clear enough: there are ways of ascribing properties to oneself on grounds that don't leave any question whether those properties actually belong to someone else.

Note that Wittgenstein is not committed saying that there is no possibility for error *at all* in the latter case; rather, he seems to be pointing to the fact that a *certain sort* of error – an error due to misidentification of the person to whom the properties should be ascribed – is not possible in these circumstances.

Now, surely the sort of self-ascription we engage in when we observe ourselves in mirrors is not *entirely* immune to this sort of error: after all, it is entirely possible that, say if I am being fooled by my trickster brother, I should turn to what I take to be a mirror, see what seems to be my reflection, and think I see myself, when really I see my brother. What this suggests, though, is simply that the characterization Wittgenstein offers in this passage isn't quite sufficient to circumscribe the phenomenon we're after: the point is not that I can't be *wrong* about who has the property in question, but that if I do make this sort of error, then I thereby lack *knowledge* that the property is had by *someone or other*. (Consider: I might seem to see a canary whenever and only whenever you do. Now suppose that I seem to see a canary – surely, unless I *know* (or at least have reason to believe) something about the way that my canary experiences are dependent on yours, *I don't know on this basis that someone sees a canary*, even though I'd be correct in forming such a belief.) the discussion from the previous section was intended to show that this is the case when we observe ourselves in mirrors: *under normal circumstances*, when one sees (what seems to be) one's reflection in (what seems to be) a mirror, the *only* person one is justified in concluding that one sees is oneself.

So let's make a distinction: call the phenomenon that Wittgenstein seems to be pointing to in this passage that of *immunity to errors OF misidentification*, and reserve the '... THROUGH misidentification' tag for the phenomenon I've been after. The relevant difference is that the latter phenomenon seems to accompany some cases that *do* 'involve the recognition of a particular person (or object)' to whom the property is ascribed, and so it is *possible* to be wrong about whom (or what) one should ascribe the property in question to. The point, though, is that when a belief is immune to error *through* misidentification in this way, one can't be wrong about this bit

and still *justified* in the ‘remainder’ (as it were) of one’s belief: one’s grounds for believing that the property is had by *something* are inseparable from his grounds for believing that it is had by *this very thing* (i.e. the thing one takes it to be had by).⁶

This recalls our attention to a crucial point, which Wittgenstein’s brief discussion doesn’t mention but which is alluded to in the definition I gave earlier: namely, that a belief will usually be immune to error through misidentification only relative to the particular *grounds* one has for holding it. (This restriction seems not to hold, though, for ‘I am in pain’ and the like, since it seems that there are no ‘indirect’ grounds on which one could come to know these sorts of things unless one also knew them by way of ‘feelings’.) I might, for instance, believe falsely that my hair is blowing in the wind because I heard someone say, ‘John’s hair is blowing in the wind’, and that person may have been talking about someone else by the same name. In this case, I still have good enough grounds for believing this about *someone* even though I’m wrong in believing it about *me*. We ran into this phenomenon, of what we might call the ‘grounds-relativity’ of this sort of immunity, earlier in discussing mirrors, and it will be important to keep it in mind through what follows.

4. *Why*, then, is it that the mirror-based beliefs we form about ourselves are immune to error through misidentification? (If you are as yet unconvinced that this is so, perhaps try replacing my ‘why’ with a ‘how’.) Given what we’ve just said, it seems reasonable to expect that this will have something to do with the way that mirrored self-recognition works: if, for instance, this involves a *judgment* identifying the person in the mirror with oneself (i.e.: a judgment to the effect that ‘*That person* (the one I see in the mirror) is F, and I am identical to that person, so I must be F’), then it seems likely that this sort of immunity probably won’t be there after all. But how else might this sort of latching-on of thoughtful reference, this ‘recognition of a particular person’, proceed?

We can begin to get a purchase on the alternative by considering patients with a cognitive disorder referred to in the scientific literature as the ‘mirror sign’ (or ‘mirrored-self

misidentification') delusion.⁷ Patients with this condition are generally able to explicitly report the presence of mirrored spaces, and can sometimes use mirrors to recognize and locate objects and persons other than themselves: crucially, however, they are entirely unable to properly regard *their own* mirrored reflections as such. Here, for example, is an exchange between Nora Breen and a patient ('FE') who suffers from the mirror sign delusion:

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.⁸

The lesson of this case should be immediate. FE appears to understand quite well what a mirror *is*; however, he is simply unable to recognize the person he sees in it as himself. But what he lacks that normal subjects have is neither knowledge nor the capacity for abstract inference:

rather, the problem is that the particular system that enables skillful mirrored self-recognition in ordinary subjects is damaged in his case. As Breen and her colleagues argue, FE's deficit seems to be at the level of *perceptual experience*: because it simply doesn't *appear* to him that the person he sees in the mirror is himself, he takes that person to be someone else who 'moves about freely' with him and family.⁹ The reason why a normal subject never makes this sort of mistake – at least in perceptually standard circumstances – is not that she makes *identity judgments* that FE doesn't make (or forms definite descriptions that FE doesn't form); indeed, normal subjects simply don't *need* to resort to identity judgments or definite descriptions when they use mirrors. Rather, it is a series of basic dispositions based immediately on a set of *practical skills* – to differentiate mirrored spaces from real ones, to recognize the face in the mirror as one's own – that do the work of fixing reference.¹⁰

This sort of idea is unfortunately missing in a lot of the scientific work that has been done on mirrored self-recognition, which is actually quite a hotly studied topic in developmental psychology and the study of animal behavior. What work in these areas seems to show is twofold: first, that only a few species – humans, chimpanzees, and orangutans, and maybe a few others – are capable of learning to use mirrors for self-observation; and second, that this skill starts to manifest itself in human children at a crucial stage in their development of a 'self-conception'. But in general, researchers in these areas regard this skill as resting on a rather complicated process of *inference*; thus here, for example, is Gordon Gallup – the initiator of this body of work – on the link between mirrored self-recognition and self-consciousness in general:

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.*¹¹

If what I have been arguing in this paper is right, then Gallup's description of mirrored self-recognition as involving an *inference* to 'the identity of the reflection' with the subject is in any

case wrong as a description of how mature subjects proceed. For what we do when we recognize ourselves in mirrors is not to bring together some ‘already existent identity’ (by which Gallup must mean ‘self-concept’) with an independent conception of ‘that object I see’: that this gets the fact wrong is shown straightforwardly by phenomenological reflection, together with the studies of patients in whom the capacity for mirrored self-recognition is impaired even while the capacity for inference – as well as the presence of conceptions of ‘myself’ and ‘the mirrored object’ – is pretty clearly there. Again, what subjects with these conditions lack is not an ability to *infer* things based on what they see; indeed, no such capacity need be operative for normal subjects when they skillfully employ mirrors as perceptual tools. Rather, their problem is simply that they *see* things in the wrong ways; they have a kind of ‘blindness’, as it were when reflecting surfaces are brought into play. And then for normal subjects in turn, it is the immediate, skillful mastery of the way that such surfaces carry information about the environment that leads them to form self-directed beliefs as they do.

5. This alternative conception of the cognitive psychology underlying mirrored self-recognition bears important affinities to the theory of self-identification that Gareth Evans develops in *The Varieties of Reference*. There, Evans argues that in lots of important cases, what makes it so that a person’s thought refers to a particular object is neither a uniquely specifying description (‘the X that is so-and-so’) nor a mere causal relation to it, but rather the thinker’s *practical mastery* of the sort of information-bearing relation (e.g. that of perception, memory, or testimony) that he stands in to that object. As Evans puts it, a subject relying on this way of thinking about things is

... in a position rather like that of the man who feels something tugging at the end of his fishing line. In such cases we are placed in a position in which we have the *practical ability* to locate the object; it is not necessary to construct some *concept* (‘the one at the

end of my line’) in order to allow the subject’s thought to reach out to its object, when he can effectively do so himself.¹²

In these sorts of cases, the subject will be immediately and non-inferentially disposed to regard information from such channels as relevant to the beliefs he forms about (and the actions he undertakes towards) the object, and it is in virtue of this *disposition* – rather than a description or a brutely causal chain – his thoughts are directed to certain things rather than others. And what I am proposing here is that in ordinary cases the process of mirrored self-recognition works in very much this way, and as a consequence the beliefs it underlies are immune to error through misidentification.

Let me close by noting two significant philosophical consequences of this theory of mirrored self-recognition. The first is that it helps to reinforce Evans’s idea that the beliefs we form about ourselves based on our immediate sensory coupling to the world around us are in certain respects no less secure than those we form based on ‘inner sense’. This adds support to a line of thought that has grown out of Strawson’s groundbreaking work on Kant (a body of work that Evans was certainly concerned to extend), namely, that it seems wrong to treat thought about oneself as a pure subject of mental states as fundamentally different from thought about oneself as a bodily being in the world. Of course this is not to say that there are *no* relevant differences: for example, self-ascription of certain mental states may be infallible and perhaps even incorrigible, and in certain cases it seems to be immune to any errors *of* misidentification at all. But the fact that my use of ‘I’ to refer to the object I see in the mirror refers as directly to me as does my use of ‘I’ to refer to the thing that thinks and entertains ideas goes a long way to showing a rather fundamental error in certain traditional approaches to personhood. To think of, and refer to, oneself as a human being is not to think of and refer to an immaterial thing – or in any case, if it is to refer to an immaterial thing then lots of our ordinary practice is massively misguided.

The second point I want to close with is closely related to this first one, and extends its consequences even further. For consider that what this paper shows to be wrong with conceiving

persons as primarily subjects of ‘inner’ mental states is not merely that we make immediate reference to ourselves as *bodies*: if the argument of this paper is right, then we also make this sort of self-reference in ways that draw immediately on the surrounded *environment*. (This is the ‘embedded’ half of the ‘embodied and embedded’ conjunct that I proposed at the start could summarize the main moral of this paper.) In other words: not only do we think of ourselves in fundamentally bodily ways, but we also think of ourselves in ways that are fundamentally ‘worldly’, in something like Heidegger’s sense of the word. Even when it comes to self-knowledge, our epistemic standing is thoroughly bound up with our capacity to make skillful use of what the environment affords.

¹ Note that I am not, or at least not immediately, concerned here with reference in speech acts. This is because the referents of linguistic items are governed by interpretive norms that do not draw as heavily on the speaker’s own grasp of Fregean *Sinne*. I am sure, though, that many elements of the present discussion could be extended to cover self-reference in speech.

² By ‘thought(s)’ I will mean something like ‘propositional attitude(s)’; the focus here is on psychological items rather than (putatively) abstract objects.

³ I should say that I don’t want too much of what I’m saying to hang on our intuitions about knowledge. The key point to note is that even if you do think that this general belief amounts to knowledge after all, it has to be admitted that it’s sort of second-rate. I think that a similar point will apply in all the other cases where I say that subjects’ beliefs fall short of knowledge.

⁴ Formally: a belief ‘ Fa ’, held by a subject S based on grounds G , is IEM with respect to the singular term ‘ a ’ iff it is not possible that (i) G does not give S sufficient grounds for knowing that Fa while (ii) G does give S sufficient grounds for knowing that $\exists xFx$. This defines what Pryor, in his paper ‘Immunity to error through misidentification’ (*Philosophical Topics* 26 (1999): 271-304), calls ‘immunity to *which-object*-misidentification’; he notes that there is another phenomenon which is sometimes regarded as a sort of IEM but which has different properties. Following Pryor, we will understand ‘giving grounds for knowing’ as ‘justifying in a way sufficient to solve the Gettier Problem’, though nothing really rests on this point.

-
- ⁵ Wittgenstein, *The Blue and Brown Books*, pp. 66-67.
- ⁶ Formally: a belief 'Fa', held by a subject *S*, is IEM* with respect to 'a' iff it is not possible that (i) *S* is wrong in believing that Fa while (ii) *S* would be correct if s/he believed that $\exists xFx$. Since 'knows' entails 'believes truly' but not vice versa, this definition makes it clear that any belief that is IEM* with respect to a singular term will also be IEM with respect to it, while the converse does not hold.
- ⁷ For a review of the literature on this condition, see K.S. Postal, '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).
- ⁸ Breen et al, 'Towards an understanding of delusions of misidentification: four case studies' (*Mind and Language* 15 (2000): 74-110), at pp. 84-85.
- ⁹ This is, indeed, the conclusion drawn by Breen et al (*op. cit.*), and also by Max Coltheart and Martin Davies in their commentary on this condition in [2000]. As Breen and her colleagues go on to note, FE suffers from a case of prosopagnosia, or 'face-blindness', which leaves him unable to recognize familiar faces and objects, and it seems that it is this condition that lies at the root of his inability to self-recognize in mirrors. The other patient they describe with this condition, TH, has intact capacities for processing faces but seems to suffer from mirror agnosia.
- ¹⁰ In this respect it is interesting to compare 'mirror sign' patients with those who suffer from schizophrenia or certain 'body schema' disruptions. Normal subjects can monitor their thoughts and the states of their bodies in information-based ways that don't depend on deliberate reflection, but patients with these conditions lose these capacities and so make self-referential judgments that seem – at least sometimes – to be subject to errors of (and perhaps even *through*) misidentification. On schizophrenia and errors of misidentification, see John Campbell, 'Schizophrenia, the space of reasons, and thinking as a motor process' (*The Monist* 82 (1999): 609-625). On disruptions of the ability to monitor one's body, see J. Cole and J. Paillard, 'Living without touch and peripheral information about body position and movement: studies with deafferented subjects' (in Bermúdez et al, eds., *The Body and the Self* (Cambridge, Mass.: MIT Press): 245-266).

¹¹ Gallup, [1977], p. 334. Quoted in Povinelli and Prince, 'When self met other' (in Ferrari and Sternberg, eds., *Self-Awareness: Its Nature and Development* (New York: The Guilford Press, 1994): 37-107), at p. 48. Emphasis added.

¹² *Varieties*, p. 172.