FINDING THE MIND IN THE BODY

Joshua Knobe

[Knobe, J. (2011). Finding the mind in the body. In Brockman, M. (ed.) *Future Science: Essays from the Cutting Edge*, New York: Random House, pp. 184–196.]

Suppose that you look out around you and see four things: a human being, a fish, a toaster, and a printer cartridge. Looking at these four things, you might be able to detect all sorts of similarities and differences, but there is one distinction here that seems especially noticeable. Some of these things have minds, while others do not. One would attribute a mind to the human being, perhaps also to the fish, but definitely not to the toaster or the printer cartridge.

This distinction comes so naturally to us that it is easy just to take it for granted, but if you stop to think about it for a moment, it begins to seem deeply puzzling. How exactly do people decide which things have minds and which do not? Most of us don't have much background in experimental psychology, cognitive neuroscience or any of the scientific disciplines that might be relevant to a question like this one. Yet, somehow if we see a fish swimming in the pond and then see a toaster popping up some toast, we immediately have the intuition that the former might have certain psychological states but that the latter most definitely does not. How might we be doing this?

One traditional answer says that the process is relatively straightforward. We figure out whether something has a mind by checking to see what it *does*. On this view, the key thing to notice about fish is that we can see them swimming around and responding to their environments in complex ways. Meanwhile, the toaster never seems to do anything interesting — all it ever does is make toast. So an obvious hypothesis would be that it is this difference in behavior that leads us to say that fish have minds while the toaster does not.

It would be hard to deny that this hypothesis has a great deal of intuitive appeal. There is something that seems deeply right about the idea that we attribute minds to objects based on their behavior, and philosophers have developed complex conceptual frameworks that spell this idea out in sophisticated detail. Such frameworks typically say that our ordinary way of thinking about the mind is something like a scientific theory. Just as a physicist might make sense of scientific observations by positing unseen entities, people make sense of each other's behavior by positing unseen mental states. The only difference is that the physicist explains his or her data in terms of purely physical factors (forces, particles, fields) while ordinary people explain behavior in terms of psychological factors (beliefs, intentions, emotions).

Yet, though the intuitive appeal of this framework is undeniable, it has recently come up against a challenge from a somewhat unexpected source. A group of people working in philosophy departments began thinking that it might be time to leave their armchairs and go out to conduct some systematic experimental studies. These 'experimental philosophers' were particularly interested in the traditional philosophical view that people's ordinary way of making sense of the world might be something like a scientific theory, and they therefore set off to put this claim to the test empirically. But, surprisingly, the experimental results did not end up conforming to the traditional view. Again and again, the results

seemed to show that people's ordinary way of making sense of the world was radically different from anything we might expect to find in a purely scientific theory.

The issues here can be quite complex, but let us focus for the moment on just one aspect of the problem. People's intuitions about whether a given entity has a mind do not appear to be based entirely on a scientific attempt to explain that entity's behavior. Instead, these intuitions seem to be influenced in a quite striking way by questions about whether that entity has the right sort of *body*.

Mind without Body

If we want to get to the bottom of these questions, a natural place to start out is by looking for an entity that has a human-like pattern of behavior but that doesn't have any kind of body. We need to find an entity that takes in information from the environment and uses this information in a complex way to attain goals. However, we also need to make sure that the entity does not have anything like a biological body in the familiar sense.

Fortunately, we can easily find an entity that meets these requirements: the modern corporation. There is a clear sense in which corporations have goals, take in information, plan accordingly. But corporations do not have biological bodies. Instead of being made up of a head, torso and limbs, they are made up of a complex hierarchy of departments and committees. So perhaps corporations can serve as a helpful case study for present purposes.

But now things begin to get interesting. The first thing to notice is that people do sometimes use sentences that seem to ascribe mental states to a corporation. For example, a person might say:

- Acme Corp. intends to release a new product in July.

Now, one might initially suppose that sentences like these are just metaphors or loose talk and that people aren't actually thinking that these corporations can literally having anything like an intention. But the experimental evidence suggests things might be a little bit more complex. In a recent neuroimaging study, I collaborated with the psychologists Anna Jenkins, David Dodell-Feder and Rebecca Saxe to look at the patterns of brain activation observed when people read sentences like these. It turns out that people reading such sentences show activation in the very same brain regions that have been traditionally associated with thinking about other minds (particularly the right temporoparietal junction, which seems to be highly selective for thinking about beliefs and intentions).

So suppose we assume for the moment that people actually are ascribing mental states to corporations. We now come to the crux of the issue. If people think that a corporation can have some kind of mind, what sort of mind do they think it can have? Do people conceive of a corporation as having the same sort of mind we might find in a human being or an animal, or do they think that there is something important that a corporation is missing? To get at these questions, I teamed up with the philosopher Jesse Prinz and ran a series of studies.

What we found was that people showed a systematic tendency to see corporations as having only one highly delimited aspect of a normal mind. In particular, people were happy to agree with sentences like these:

- Acme Corporation believes that its profit margin will soon increase.
- Acme Corporation intends to release a new product this January.

- Acme Corporation wants to change its corporate image.

But suppose we then shift over to sentences that ascribe to a corporation some kind of feeling or experience. For example:

- Acme Corporation is now experiencing great joy.
- Acme Corporation is getting depressed.
- Acme Corporation is feeling upset.

People regarded these sentences as completely wrong, sometimes laughably so. In other words, people seem to think that corporations are capable of deciding, intending, knowing, and so on, but that they are not capable of truly feeling or experiencing anything. In the jargon of philosophy, corporations are regarded as utterly lacking in *phenomenal consciousness*.

Here again, we seem to be faced with a result that initially seems perfectly natural and obvious but begins to look more and more puzzling as one examines it further. Why exactly can't a corporation feel upset? A corporation can certainly *behave* in ways that would be characteristic of feeling upset, yet one somehow gets the sense that the corporation wouldn't truly be *feeling* anything at all. How do we arrive at this intuition? One possible answer is that people don't think that corporations can have consciousness because corporations don't have the right sorts of bodies.

Minds and Machines

But, of course, we will never be able to isolate the role of the body if we just restrict our attention to corporations. Corporations differ from human beings in numerous respects, and any of these differences could be explaining the observed effects. What we really need, then, is an entity that is almost exactly the same as a human being except that it lacks a biological body.

The philosopher Bryce Huebner came up with the perfect way to fulfill these requirements. He conducted a series of studies in which participants were asked to imagine a robot that has been designed to act like a person. Although the robot is presumably made of silicon and metal, it is described as behaving exactly like a human being on all possible psychological tests. The question now is what sorts of mental states people will be willing to ascribe to it.

Strikingly, the answer is that people ascribe to the robot exactly the same sorts of states that they are willing to ascribe to a corporation. They are happy to say:

- It believes that triangles have three sides.

But they are unwilling to say:

- It feels happy when it gets what it wants.

In other words, what we see arising here is exactly the same asymmetry we observed for intuitions about corporations. Once again, people are describing an entity without a biological body as having an ability to have states like beliefs but not as having a capacity for genuine feeling or experience.

But notice what is happening this time. The robot is described as behaving *exactly* like a human being in all situations. So any difference between the mental states we ascribe to it and the mental states we ascribe to a human being can't be understood in terms of an attempt to predict behavior. It

must be that the body is playing some role here. Something about the presence of our faces, our flesh, our biological nature must be triggering people to think that we have phenomenal consciousness.

Mind and Flesh

Thus far, we have been considering cases in which an entity is seen as not having a body. But suppose we now go in the opposite direction. Suppose we find a case in which an entity is seen as *especially* embodied, a case in which we associate this entity with a body even more than we would in cases of ordinary human interaction. What sort of mind would they ascribe in a case like that?

I was discussing this question one day with the psychologist Kurt Gray when someone happened to overhear us and suggested an interesting new approach. It turned out that there was a recently published photography book in which each model was depicted in different stages of undress. This fact alone may seem a bit unsurprising, but in this particular case, the book was composed in a particularly systematic way. For each of the models, there was a completely clothed photograph and then, on the facing page, a photograph of the same person, in the same position, with the same facial expression, only this time completely naked. It was a cognitive scientist's dream, the artistic equivalent of a perfectly controlled study. We immediately decided to go ahead and run a new experiment, with the pictures from this book as stimulus materials.

In collaboration with our colleagues Paul Bloom, Mark Sheskin and Lisa Feldman Barrett, we put together the experimental design. Each participant would receive a photograph and then would be asked to guess, just on the basis of that one photograph, how much the model depicted was capable of having various different mental states. So a participant might be asked:

- Compared to the average person, how much is Erin capable of self-control?
- Compared to the average person, how much is Erin capable of feeling fear?
- Compared to the average person, how much is Erin capable of planning?

But, of course, the stimuli were designed in such a way as to let us look systematically at the relationships between different variables. The questions asked both about ordinary non-phenomenal states (self-control, planning) and about states that involved phenomenal consciousness (feeling fear, feeling pleasure). And each participant was randomly assigned to receive either a picture of a clothed model or a picture of a naked model.

Before I tell you the results, let's take just a moment to consider the different predictions one might make in a case like this. One obvious prediction would be that showing the model naked would make her vulnerable to a kind of 'objectification.' Participants might come to think of her more as a physical object, a mere thing, and they might therefore be less inclined to see her as having mental states. But then again, we can also imagine another sort of prediction, going in a quite different direction. Perhaps there is something about an awareness of the body that makes people more inclined to ascribe phenomenal consciousness. So the effect might actually end up going the opposite way. A focus on the body could make participants *more* inclined to think in terms of feelings and experiences.

With these predictions in mind, we can now turn to the actual results. For the non-phenomenal states, we ended up finding exactly what one would expect. The more salient a person's body was made, the less inclined participants were to ascribe these states. In other words, if you want people to take you as someone capable of complex planning and self-control, your best bet is not to have these

people looking at pictures of you naked. No surprises there. This is exactly the point that has been made repeatedly, and with great sophistication, in existing work within feminist theory.

But now comes the surprising part. For the phenomenal states, we did not find this same pattern. In fact, we found just the opposite. When the model was depicted naked, people were actually more inclined to think that she was capable of having feelings and experiences. They were more inclined to think she was capable of feeling fear, more inclined to think she was capable of feeling pleasure. In fact, on all the different measures we used, we always found that making the body more salient made people more inclined to ascribe feelings.

But the effect does not stop there. Gray has tested this same basic hypothesis using a whole series of imaginative experimental techniques. He has given participants information about a person's blood type, asked them to judge a person's physical attractiveness, even shown them pornographic images. Always, the result is the same. The more one makes participants focus on the body, the more they tend to ascribe feelings and experiences.

All in all, then, it does not appear that these phenomena are best understood in terms of a notion of 'objectification.' It is not as though participants are coming to think of a person as being a mere *object,* like a toaster or a printer cartridge. Rather, what we see emerging is a more complex pattern. Participants are thinking of the person as having less of one part of the mind but more of another. So perhaps it would be better to say that a focus on the body leaves us thinking of the person as an *animal*. That is, it leads us to think of the person as having more of the part of the mind we associate with animals (fear, pleasure, pain) and less of the part we regard as distinctively human (complex reasoning, planning, self-control).

Perceiving the Mind

With all this experimental evidence on the table, we can now return to our original question. We wanted to get a better understanding of the process people use to figure out which entities have minds and which do not. So what exactly is this experimental evidence telling us?

The answer may come as a surprise. The key message coming out of the experimental evidence seems to be that the whole question was a mistaken one. It is beginning to look like we might have been wrong to go searching for something like 'the process people use to figure out which entities have minds.' The trouble is that there just doesn't seem to be any single unified process that fits the bill. Instead, there appear to be two distinct processes here – one for figuring out whether an entity is capable of having states like beliefs and goals, another for figuring out whether an entity is capable of genuine feelings. If we really want to get to the bottom of these issues, we will have to address each of these processes separately.

But when the question is reformulated in this way, a new answer becomes possible. We might discover that people's understanding of beliefs and goals involves something like a quasi-scientific attempt to explain human behavior, but we should not immediately assume that the same holds for people's understanding of feelings and experiences. On the contrary, all the evidence suggests that this latter process is deeply different. Our ordinary attributions of phenomenal consciousness do not appear to be based entirely on behavior; they do not appear to be purely scientific; they do not seem to serve primarily to aid prediction or explanation. Above all, they seem to be wrapped up in some fundamental way with an awareness of the body.