

The psychological theory of persons

Last week we were discussing dualist views of persons, according to which human beings are immaterial things distinct from their bodies. We closed by discussing some problems for that view.

One natural sort of response to these problems is to move from dualism, the view that we are immaterial things, to materialism, the view that we are material things. If asked what sort of material thing we could be, there seems at first only one plausible answer: we are our bodies.

But this view too faces some serious problems. One comes up in the following passage from the dialogue we read for class today:

MILLER: Okay. Now it seems to me we are even able to imagine awakening and finding ourselves to have a *different* body than the one we had before. Suppose yourself just as I have described you. And now suppose you finally open your eyes and see, not the body you have grown so familiar with over the years, but one of a fundamentally different shape and size.

WEIROB: Well, I should suppose I had been asleep for a very long time and lost a lot of weight—perhaps I was in a coma for a year or so.

MILLER: But isn't it at least conceivable that it should not be your old body at all? I seem to be able to imagine awakening with a totally new body.

WEIROB: And how would you suppose that this came about?

MILLER: That's beside the point. I'm not saying I can imagine a procedure that would bring this about. I'm saying I can imagine it happening to me. In Kafka's *Metamorphoses*, someone awakens as a cockroach. I can't imagine what would make this happen to me or anyone else, but I can imagine awakening with the body of a cockroach. It is incredible that it should happen—that I do not deny. I simply mean I can imagine experiencing it. It doesn't seem contradictory or incoherent, simply unlikely and inexplicable.

The argument that Miller is making here is closely related to Descartes' argument for dualism.

Descartes argued that it is possible that I exist even though no material things exist, and hence that it is possible that I exist without a body; he then argued that from the possibility of me existing without a body, it follows that I am not identical to my body (or any other).

Miller is not imagining existing without any body; he is imagining existing in a *different* body. He claims that this is a possible scenario: it is, as says, quite unlikely, and it is hard to see how this sort of scenario could come to pass, but it is not incoherent or impossible. We should ask two questions about this argument:

1. Is Miller right that it is possible for me to exist in a different body?
2. Is Miller right that if this is possible, I cannot be identical to my body?

We will return to the materialist view of personal identity next week. But for now, the thing to note is that we seem to be in trouble. We have two views of persons on the table -- according to the first, I am an immaterial thing, and according to the second, I am a material thing. Both seem to face serious problems; but it seems as though one or the other must be right. After all, there are such things as persons, and what could they be if not either material or immaterial things?

In the passage we read, Miller argues in effect that this dilemma overlooks a third view about persons, which is expressed in the following passage:

WEIROB: With rivers and baseball games, I can see that they are made up of parts connected in a certain way. The connection is, of course, different in the two cases, as is the sort of "part" involved. River parts must be connected physically with other river parts to form a continuous whole. Baseball innings must be connected so that the score, batting order, and the like are carried over from the earlier inning to the later one according to the rules. Is there something analagous we are to say about persons?

MILLER: Writers who concern themselves with this speak of "person-stages." That is just a stretch of consciousness, such as you and I are aware of now. I am aware of a flow of thoughts and feelings that are mine, you are aware of yours. A person is just a whole composed of such stretches as parts, not some substance that underlies them, as I thought yesterday, and not the body in which they occur, as you seem to think. That is the conception of a person I wish to defend today.

WEIROB: So when I awoke and said to myself, "I am the one who was so rude to Sam Miller last night," I was judging that a certain stretch of consciousness I was then aware of, and an earlier one I remembered having been aware of, form a single whole of the appropriate sort—a single stream of consciousness, we might say.

MILLER: Yes, that's it exactly. You need not worry about whether the same immaterial soul is involved, or even whether that makes sense. Nor need you worry about whether the same body is involved, as indeed you do not since you don't even have to open your eyes and look. Identity is not, so to speak, something under the person-stages, nor in something they are attached to, but something you build from them.

On the view of persons Miller is sketching, we are supposed to understand persons as collections of *person-stages*. You can think of a person-stage as being just a person at a single time.

So the theory of persons that Miller is sketching is a theory according to which persons are a series of person-stages connected *in the right way*. So what we need to figure out is: what does it mean for a series of person-stages to be connected in the right way?

Miller's answer to this question involves memory. (The theory he is giving here was first defended by the 17th century English philosopher John Locke.) We can put it like this:

The memory theory of persons

Two person-stages are stages of the same person if and only if the later person-stage contains memories of the earlier person-stage.

Weirob raises a problem for this sort of theory:

But we must distinguish—as I'm sure you will agree—between *actually* remembering and merely *seeming* to remember. Many men who think that they are Napoleon claim to remember losing the battle of Waterloo. We may suppose them to be sincere, and to really seem to remember it. But they do not actually remember because they were not at the battle and are not Napoleon.

Why is this a problem for the memory theory?

We can distinguish two replies to this objection:

1. We stipulate that for one person-stage to contain memories of another, the memories must be *real memories*, not false ones. And for memories to be real, they have to be memories of the same person.
2. We stipulate that for one person-stage to contain memories of another, the memories have to be caused in the right way by the experiences that they seem to be memories of.

Which of these replies is more plausible?

So that's one sort of problem which arises from having extra, false memories; but other problem arise in cases in which we *lack* memories. John Locke raised the problem in this way:

§ 22. But is not a Man Drunk and Sober the same Person, why else is he punish'd for the Fact he commits when Drunk, though he be never afterwards conscious of it? Just as much the same Person, as a Man that walks, and does other things in his sleep, is the same Person, and is answerable for any mischief he shall do in it.]

What problem does Locke have in mind here? What does the memory theory say about cases of drunkenness or sleep?

A related problem was raised by Thomas Reid, a Scottish contemporary of Locke:

Suppose a brave officer to have been flogged when a boy at school, for robbing an orchard, to have taken a standard from the enemy in his first campaign, and to have been made a general in advanced life: Suppose also, which must be admitted to be possible, that, when he took the standard, he was conscious of his having been flogged at school, and that when made a general he was conscious of his taking the standard, but had absolutely lost the consciousness of his flogging.

These things being supposed, it follows, from Mr Locke's doctrine, that he who was flogged at school is the same person who took the standard, and that he who took the standard is the same person who was made a general. Whence it follows, if there be any truth in logic, that the general is the same person with him who was flogged at school. But the general's consciousness does not reach so far back as his flogging—therefore, according to Mr Locke's doctrine, he is not the person who was flogged. Therefore, the general is, and at the same time is not the same person with him who was flogged at school. . . .⁸

What objection does Reid have in mind here?

One way to put Reid's objection using the language of "person-stages" goes like this.

Let A be a person stage of yours at the age of 5, and let B be a person stage of yours at the age of 13, and let C be a person stage of yours at the age of 17. Then then the following seems possible:

C has memories of the experiences of B, and B has memories of the experiences of A,
but C does not have memories of the experiences of A.

The problem is that this seems to leads to the following argument:

1. Two person stages are the same person if and only if if the later has memories of the earlier.
(The Memory Theory)
2. C has memories of the experiences of B.
3. $C=B$ (1,2)
4. B has memories of the experiences of A.
5. $B=A$ (1,4)
6. C does not have memories of the experiences of A.
7. $C \neq A$ (1,6)
8. $C=A$ (3,5)

C. $C=A$ & $C \neq A$ (7,8)

What sort of argument is this? Is the argument designed to convince you that the conclusion is true? If not, what is the point of the argument?

Reid's argument seems to show that premise (1) of the argument just sketched is false. That was:

1. Two person stages are the same person if and only if if the later has memories of the earlier.

This premise can be thought of as having two parts:

- 1a. If two person stages are the same person, then the later has memories of the earlier.
- 1b. If a later person-stage has memories of the earlier one, then they are the same person.

Premise (1) is equivalent to the combination of (1a) and (1b). We mentioned earlier that if-then sentences are sometimes called *conditionals*. Sentences like premise (1) are called *biconditionals*, because they are like the combination of two conditionals, 'if p, then q' and 'if q, then p.' You can write this as 'p if and only if q' (or 'q if and only if p' -- order does not matter).

One suggestion for the memory theory is that they should give up (1a). This would be enough to solve the problem posed by the example of drunkenness and to block Reid's argument.

Does this really solve the problem posed by Reid's argument? Suppose we replace premise (1)

1. Two person stages are the same person if and only if if the later has memories of the earlier.

with premise (1b):

- 1b. If a later person-stage has memories of the earlier one, then they are the same person.

Then we would have the following argument:

- 1b. If a later person-stage has memories of the earlier one, then they are the same person. (The Modified Memory Theory)
2. C has memories of the experiences of B.
3. $C=B$ (1b,2)
4. B has memories of the experiences of A.
5. $B=A$ (1b,4)
6. C does not have memories of the experiences of A.
7. $C \neq A$ (1b,6)
8. $C=A$ (3,5)

C. $C=A$ & $C \neq A$ (7,8)

Is this argument still valid? If not, why not?

The argument is no longer valid, because the inference from premises (1b) and (6) to (7) is invalid. You can think of that inference as being of the form 'If p, then q; not-p; therefore, not-q.' But this is not a valid form of inference.

Let's put all these points together. First, in order to handle the case of false memories, the memory theory has to incorporate some distinction between memories which are caused in the right way, and ones which are not.

Second, in order to respond to Reid's argument, the memory theorist should reject the claim that

If two person stages are the same person, then the later has memories of the earlier.

That means that they have to reject our theory as initially formulated, which was

The memory theory of persons (1st version)

Two person-stages are stages of the same person if and only if the later person-stage contains memories of the earlier person-stage.

Instead, the memory theory should be stated as follows:

The memory theory of persons (2nd version)

Two person-stages are stages of the same person if the later person-stage contains memories, caused in the right way, of the earlier person-stage.

Then we can add this claim:

If A and B are the same person, and B and C are the same person, then A and C are the same person.

Our final version of the memory theory might then intuitively be stated like this: A and B are the same person if and only if either (1) A contains memories, caused in the right way, of B, or (2) there is some series of person-stages connecting A and B which is such that each person-stage in the series has memories (caused in the right way) of the immediately preceding person-stage in the series.

Does this solve all the problems with the memory theory that we have discussed so far?

