

Materialism &
the
psychological
theory of
persons

Last week we focused on the central argument for the conclusion that persons are not material objects, and some problems which arise for making sense of the idea that immaterial minds could interact with a material world.

The problems involved with making sense of causal connections between immaterial minds and a material world might lead one to pursue a view according to which persons are material, rather than immaterial, things. That is, one might want to explore the following view:

Materialism about human beings

We are material (physical) objects.

Remember that last week I introduced three basic world views: materialism, idealism, and dualism. Materialism is the view that everything is a physical object - everything is composed of the sorts of things studied in physics - protons, electrons, etc. - and nothing else. Since materialism is materialism about **everything**, if this overall world-view is going to be correct, we had better be able to develop a materialist theory of human beings.

One strength of this sort of view should be clear: it avoids the problems, faced by the dualist, with explaining how persons could interact with the physical world. If materialism is true, there is no more puzzle about how human beings could interact with the physical world than there is about how two billiard balls could bounce off of each other.

But this view also faces some problems. These can be divided into two categories: problems to do with the existence of material objects in general, and problems about the idea that human beings in particular are material objects.

The first sort of problem is exemplified by the story of the ship of Theseus, which is discussed in the excerpt from van Inwagen.

Materialism about human beings

We are material (physical) objects.

The first sort of problem is exemplified by the story of the ship of Theseus, which is discussed in the excerpt from van Inwagen.

Let us consider the famous story (famous among metaphysicians, anyway) of the Ship of Theseus. The hero Theseus has a ship, which is entirely composed of wooden planks. Very gradually, over the course of years, the planks are removed from the ship and replaced. The replacement is so gradual that Theseus and his crew are able to be almost continuously at sea, engaged the while in a long series of adventures with a nautical setting. The planks that have been removed from the ship are not destroyed but are rather stacked in a certain field. When all of the original planks have been replaced, Stilpo the shipwright notices that the field contains all of the components needed to build a ship. Stilpo puts the planks together and puts them together in such a way that they are arranged exactly as they were when they composed Theseus's ship on the day he first took command of it. Stilpo takes his new ship to sea for a shakedown cruise, and his ship and Theseus's ship pass each other at sea.

Let's introduce some names which will help us to talk about this story clearly.

Original Ship = the material object on which Theseus sets forth on the first day of our story.

Reconstructed Ship = the material object on which Stilpo is sailing when he passes Theseus.

Continuous Ship = the material object on which Theseus is sailing when he passes the Reconstructed Ship, piloted by Stilpo.

Let us consider the famous story (famous among metaphysicians, anyway) of the Ship of Theseus. The hero Theseus has a ship, which is entirely composed of wooden planks. Very gradually, over the course of years, the planks are removed from the ship and replaced. The replacement is so gradual that Theseus and his crew are able to be almost continuously at sea, engaged the while in a long series of adventures with a nautical setting. The planks that have been removed from the ship are not destroyed but are rather stacked in a certain field. When all of the original planks have been replaced, Stilpo the shipwright notices that the field contains all of the components needed to build a ship. Stilpo puts the planks together and puts them together in such a way that they are arranged exactly as they were when they composed Theseus's ship on the day he first took command of it. Stilpo takes his new ship to sea for a shakedown cruise, and his ship and Theseus's ship pass each other at sea.

Materialism about human beings

We are material (physical) objects.

Original Ship = the material object on which Theseus sets forth on the first day of our story.

Reconstructed Ship = the material object on which Stilpo is sailing when he passes Theseus.

Continuous Ship = the material object on which Theseus is sailing when he passes the Reconstructed Ship, piloted by Stilpo.

What we want to know is: what is the relationship between the material objects named by these three names?

At least one thing is clear. Since a ship cannot pass itself at sea, we know that

Reconstructed Ship \neq Continuous Ship

How about Original Ship? One might think that the following principle seems pretty plausible:

If x and y are material things which have exactly the same parts, then $x=y$.

But if this principle is true, we know something else:

Reconstructed Ship = Original Ship

Let us consider the famous story (famous among metaphysicians, anyway) of the Ship of Theseus. The hero Theseus has a ship, which is entirely composed of wooden planks. Very gradually, over the course of years, the planks are removed from the ship and replaced. The replacement is so gradual that Theseus and his crew are able to be almost continuously at sea, engaged the while in a long series of adventures with a nautical setting. The planks that have been removed from the ship are not destroyed but are rather stacked in a certain field. When all of the original planks have been replaced, Stilpo the shipwright notices that the field contains all of the components needed to build a ship. Stilpo puts the planks together and puts them together in such a way that they are arranged exactly as they were when they composed Theseus's ship on the day he first took command of it. Stilpo takes his new ship to sea for a shakedown cruise, and his ship and Theseus's ship pass each other at sea.

Materialism about human beings

We are material (physical) objects.

Original Ship = the material object on which Theseus sets forth on the first day of our story.

Reconstructed Ship = the material object on which Stilpo is sailing when he passes Theseus.

Continuous Ship = the material object on which Theseus is sailing when he passes the Reconstructed Ship, piloted by Stilpo.

Reconstructed Ship \neq Continuous Ship

Reconstructed Ship = Original Ship

But once we know these two things, we know something else:

Original Ship \neq Continuous Ship

After all, identity is transitive: if $x=y$ and $y=z$, then it follows that $x=z$. Hence if $x=y$ and $x\neq z$, we know that $y\neq z$.

Let us consider the famous story (famous among metaphysicians, anyway) of the Ship of Theseus. The hero Theseus has a ship, which is entirely composed of wooden planks. Very gradually, over the course of years, the planks are removed from the ship and replaced. The replacement is so gradual that Theseus and his crew are able to be almost continuously at sea, engaged the while in a long series of adventures with a nautical setting. The planks that have been removed from the ship are not destroyed but are rather stacked in a certain field. When all of the original planks have been replaced, Stilpo the shipwright notices that the field contains all of the components needed to build a ship. Stilpo puts the planks together and puts them together in such a way that they are arranged exactly as they were when they composed Theseus's ship on the day he first took command of it. Stilpo takes his new ship to sea for a shakedown cruise, and his ship and Theseus's ship pass each other at sea.

Materialism about human beings

We are material (physical) objects.

Reconstructed Ship \neq Continuous Ship

Reconstructed Ship = Original Ship

Original Ship \neq Continuous Ship

So we have established three conclusions about the objects in our story. For our purposes, the most interesting of these conclusions is the last one: the claim that original ship is an object distinct from continuous ship.

This claim is somewhat puzzling. After all, it would not seem to Theseus (unless he was a philosopher) that he had changed ships. One wonders: what about the situation explains the fact that Continuous ship is not the same object as Original ship?

An initially tempting answer is that this has something to do with Stilpo, and the reconstructed ship. But on reflection this does not look so plausible. It does not seem plausible to say that whether or not Theseus is aboard the same material object can depend on what Stilpo is doing in the shipyard, hundreds of miles away.

Let us consider the famous story (famous among metaphysicians, anyway) of the Ship of Theseus. The hero Theseus has a ship, which is entirely composed of wooden planks. Very gradually, over the course of years, the planks are removed from the ship and replaced. The replacement is so gradual that Theseus and his crew are able to be almost continuously at sea, engaged the while in a long series of adventures with a nautical setting. The planks that have been removed from the ship are not destroyed but are rather stacked in a certain field. When all of the original planks have been replaced, Stilpo the shipwright notices that the field contains all of the components needed to build a ship. Stilpo puts the planks together and puts them together in such a way that they are arranged exactly as they were when they composed Theseus's ship on the day he first took command of it. Stilpo takes his new ship to sea for a shakedown cruise, and his ship and Theseus's ship pass each other at sea.

Materialism about human beings

We are material (physical) objects.

Reconstructed Ship \neq Continuous Ship

Reconstructed Ship = Original Ship

Original Ship \neq Continuous Ship

This claim is somewhat puzzling. After all, it would not seem to Theseus (unless he was a philosopher) that he had changed ships. One wonders: what about the situation explains the fact that Continuous ship is not the same object as Original ship?

A more promising answer to our question is: Original Ship and Continuous Ship are distinct because they are not composed of the same parts. This, after all, was what allowed us to imagine Stilpo and his Reconstructed Ship. (If the Original Ship had never lost some of its parts, there would have been nothing to make the Reconstructed Ship out of.) This suggests the following principle:

Essentiality of parts

If x and y are material objects, and x and y have different parts, then $x \neq y$.

Essentiality of parts

If x and y are material objects, and x and y have different parts, then $x \neq y$.

Materialism about human beings

We are material (physical) objects.

Let's now return to the topic of the nature of human persons, and ask how well Materialism fares when combined with the principle of the essentiality of parts, which seemed to be established by the story of the Ship of Theseus.

As van Inwagen points out, these two theses do not seem to fit together very well:

If I am, as the physicalists say, a living organism or a part of one, then I have "lost" almost all of the atoms that composed me ten years ago and I am now made almost entirely of atoms that existed ten years ago but were then parts of other things or parts of nothing at all. It is true that I have the same brain-cells I had ten years ago (minus those that have died in the interval), but each of those brain-cells is now made of atoms that were not parts of it ten years ago.

If I am a physical thing, therefore, I am made of different matter from the matter that composed the physical thing that bore the name 'Peter van Inwagen' ten years ago. The physicalist is forced to say that all of our statements that imply that I existed ten years ago must be, strictly speaking, false.

What is the problem here?

So far we have been thinking of the materialist as committed to the idea that we are identical to our bodies. What the example of the Ship of Theseus seems to show is that if this sort of materialist wants to say that I am the same person as the person who was called "Jeff Speaks" 10 years ago, then he must reject the thesis of the essentiality of parts. That means that he must give a different interpretation of the story of the Ship of Theseus than the one we gave.

Materialism about human beings

We are material (physical) objects.

The Ship of Theseus poses a general problem for our understanding of how material objects can exist over a period of time during which they lose some of their parts. In the selection from Locke, though, we also get an argument against the idea that human beings, in particular, are identical to their bodies.

This argument is based on Locke's example of the prince and the cobbler.

For should the Soul of a Prince, carrying with it the consciousness of the Prince's past Life, enter and inform the Body of a Cobbler as soon as deserted by his own Soul, every one sees, he would be the same Person with the Prince, accountable only for the Prince's Actions

What sort of example is Locke imagining here?

This seems to be a problem for the simple materialist views of human persons introduced above. If Locke is right, and we can coherently imagine cases in which two persons "swap bodies", then it seems that we cannot be identical to our bodies.

For should the Soul of a Prince, carrying with it the consciousness of the Prince's past Life, enter and inform the Body of a Coffer as soon as deserted by his own Soul, every one sees, he would be the same Person with the Prince, accountable only for the Prince's Actions

The sort of case Locke imagines is not just a problem for the simple materialist theories of persons discussed above; it also suggests another theory of the nature of persons. Why, in this sort of case, do we all think that the person corresponding to the cobbler-body would be the prince? The key seems to be the fact that this person would have the "consciousness of the Prince's past life."

This suggests that what is essential to persons is neither an immaterial soul nor a material body, but rather some sort of **continuity of consciousness**. This is the central idea of a third major theory theory of the nature of persons (the first two being the view that we are immaterial souls and that we are our bodies), which is sometimes called the **psychological theory** or the **memory theory** of personal identity. Locke is usually regarded as the first to defend this sort of theory.

This theory might be expressed as follows:

The memory theory of persons

If x and y are persons, then $x=y$ if and only if x has memories of y (or vice versa).

Though this theory is obviously quite different than the sorts of materialist theories with which we began, one might well think that this is the sort of theory a materialist could endorse. After all, this theory does **not** postulate immaterial souls, or indeed anything which is not a material thing. The theory just says that a given person might be a different material thing at different times.

This suggests that what is essential to persons is neither an immaterial soul nor a material body, but rather some sort of **continuity of consciousness**. This is the central idea of a third major theory of the nature of persons, which is sometimes called the **psychological theory** or the **memory theory** of personal identity. Locke is usually regarded as the first to defend this sort of theory.

This theory might be expressed as follows:

The memory theory of persons

If x and y are persons, then $x=y$ if and only if x has memories of y (or vice versa).

Though this theory is obviously quite different than the sorts of materialist theories with which we began, one might well think that this is the sort of theory a materialist could endorse. After all, this theory does **not** postulate immaterial souls, or indeed anything which is not a material thing. The theory just says that a given person might be a different material thing at different times.

As Locke was aware, this theory has some surprising consequences. Here is one sort of problem that Locke raised for his own theory:

§ 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 is the problem here? How should Locke respond?

The memory theory of persons

If x and y are persons, then $x=y$ if and only if x has memories of y (or vice versa).

But the problems faced by the memory theory go well beyond these sorts of surprising consequences. As Thomas Reid, a Scottish contemporary of Locke, argued, certain sorts of examples seem to show that the theory is self-contradictory.

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.



The memory theory of persons

If x and y are persons, then $x=y$ if and only if x has memories of y (or vice versa).

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.

As with the example of the Ship of Theseus, it will be useful to introduce some names to bring out the sort of example Reid has in mind.

A = the boy at the time of the flogging
B = the officer at the time of the standard-taking
C = the general in "advanced life"

Then what Reid seems to be saying is that the following sort of scenario is 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.

We can see why this sort of scenario is problematic for the memory theory by laying out the following *reductio* argument against that theory:

Reid's reductio of the memory theory

1. x and y are the same person if and only 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)

Reid's reductio of the memory theory

1. x and y are the same person if and only 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≠A (1,6)
 8. C=A (3,5)
-
- C. C=A & C≠A (7,8)

Reid's argument is a powerful one. It assumes only the transitivity of identity and the possibility of the sort of scenario described above. It is extremely difficult to deny that such scenarios are, in fact, possible.

So let's suppose that Reid's argument shows that the memory theory of persons, as stated above, is false. To respond to this argument, then, it seems that a proponent of that theory should try to find a way to reformulate her theory in such a way that it avoids Reid's objection.

The memory theory of persons

If x and y are persons, then $x=y$ if and only if x has memories of y (or vice versa).

The problem for the Memory Theory is, pretty clearly, the derivation of Premise 7 in the above argument, which says that the general is not the same person as the boy. Somehow, the memory theorist must deny one of the premises used to this point.

One way to see how this might work is to think of the memory theory as breaking down into two parts:

The memory requirement

If A does **not** remember any of the experiences of B, then $A \neq B$.

The memory guarantee

If A does remember an experience of B, then $A=B$.

To get to premise 7, the crucial assumption is the **memory requirement**. The idea there is that lack of memories ensures non-identity - which is what the memory requirement says.

The memory theory of persons

If x and y are persons, then $x=y$ if and only if x has memories of y (or vice versa).

The problem for the Memory Theory is, pretty clearly, the derivation of Premise 7 in the above argument, which says that the general is not the same person as the boy. Somehow, the memory theorist must deny one of the premises used to this point.

One way to see how this might work is to think of the memory theory as breaking down into two parts:

The memory requirement

If A does **not** remember any of the experiences of B, then $A \neq B$.

The memory guarantee

If A does remember an experience of B, then $A=B$.

To get to premise 7, the crucial assumption is the **memory requirement**. The idea there is that lack of memories ensures non-identity - which is what the memory requirement says.

A natural thought for the memory theorist to have is, therefore, that perhaps the memory requirement should be abandoned, and that a good psychological theory of persons should stick with the memory guarantee alone.

However, this claim does not tell us everything we might want a theory of personal identity to tell us. It does not, in particular, seem to tell us exactly when x and y are the same person; if x does not have memories of y and y does not have memories of x , then this claim is simply silent on the question of whether x is y . But this might not seem very satisfactory; shouldn't a theory of persons explain what it takes, in any case, for x to be the same person as y ?

We can resolve this problem by reminding ourselves that identity is transitive: **If A and B are the same person, and B and C are the same person, then A and C are the same person.**

But, given the transitivity of identity, we know from the above claim that even if x does not have memories of y , x **must** be the same person as y if there is someone of whom x has memories and that person also shares memories with y .

And this suggests a way in which the memory requirement might be modified so as to avoid the problems raised by Reid's example.

The memory requirement

If A does **not** remember any of the experiences of B, then $A \neq B$.

The memory guarantee

If A does remember an experience of B, then $A=B$.

However, this claim does not tell us everything we might want a theory of personal identity to tell us. It does not, in particular, seem to tell us exactly when x and y are the same person; if x does not have memories of y and y does not have memories of x, then this claim is simply silent on the question of whether x is y. But this might not seem very satisfactory; shouldn't a theory of persons explain what it takes, in any case, for x to be the same person as y?

We can resolve this problem by reminding ourselves that identity is transitive: **If A and B are the same person, and B and C are the same person, then A and C are the same person.**

But, given the transitivity of identity, we know from the above claim that even if x does not have memories of y, x **must** be the same person as y if there is someone of whom x has memories and that person also shares memories with y.

And this suggests a way in which the memory requirement might be modified so as to avoid the problems raised by Reid's example.

Perhaps rather than saying that if $A=B$ then there must be a **direct** memory relation between A and B - A remembering one of B's experiences - we should weaken the requirement to say that if $A=B$ then there must be at least an **indirect** memory relation between A and B.

For there to be an indirect memory relation between A and B is for there to be a series of persons connecting A and B which is such that **each person in the series has memories of the immediately preceding person in the series, and both A and B are members of the series.**

With this understanding of an indirect memory relation in hand, we can see that there is an indirect memory relation between the general and the boy, since there is a series of persons connecting the two such that each member of the series has memories of the one just before it.

Hence Reid's objection to the memory theory can be avoided if we replace the memory requirement with the following:

The modified memory requirement

If there is no memory relation between A and B, whether direct or indirect, then $A \neq B$.

The modified memory requirement

If there is no memory relation between A and B, whether direct or indirect, then $A \neq B$.

The memory guarantee

If A does remember an experience of B, then $A=B$.

For all that we have said so far, then, it seems that if we think of the psychological theory of persons as involving the following two claims - the memory guarantee + the modified memory requirement - then it seems that the psychological theory of persons offers the materialist a promising theory of personal identity.

Next time we will continue our discussion of this theory. As we will see, some problems which arise in connection with this view call into question not just the nature of persons, but their reality.