

1 Personal Identity

It is a truism that people survive through change. When I was 14, I took a trip to Alaska and caught a 25lb. King salmon. Today, quite a lot about me has changed; I'm taller and a bit heavier, my taste in music and political views are different, and I'm in the most densely populated state as opposed to the least. But it is still me who caught that fish in Alaska. The 14 year old individual in Alaska and the individual in New Jersey today are *one and the same person*.¹

This seems perfectly obvious, but when we recognize that so much about who I am has changed through time, we are led to ask what it is in virtue of which my past and current selves are one person? That is, we are led to investigate the **identity conditions for persons**. We want to fill out the following schema in such a way that the conditions we list in the blank on the right provide a real definition² for the notion on the left.

A is the same person as B iff³ _____

Since what we are after is personal *identity* we can abbreviate "A is the same person as B" with " $A = B$ ". The identity relation has some important properties. It is **reflexive** ($A=A$, for all A), **symmetric** (if $A=B$, then $B=A$), and **transitive** (if $A=B$ and $B=C$, then $A=C$). So, whatever condition we come up with to fill in the blank must have these properties as well.

1.1 The body theory

A good place to start is with the means we use to recognize a person on different occasions. Generally, we recognize that a person is the same person we met earlier based on their bodily features. So, we might try the following definition:

A is the same person as B iff A and B share the same body.

But this proposal runs into obvious difficulties. First, when someone loses an arm, they no longer possess the same body they used to, but we don't want to say that they are a different person. Even if you think that in some sense the person who loses an arm still has the same body, it so happens that every 7 years or so, every cell in your body is replaced, so every 7 years, you get a whole new body. But we don't become a new person every 7 years.

What these examples indicate is that it isn't the **material composition** of our bodies that is important. Maybe what's important is that there be some sort of **continuity** between a person's body at one time and their body at another. As a second pass, we might try:

A is the same person as B iff A's body is causally continuous with B's body.

This proposal gets around the earlier worries, but it runs into the counter-example made famous by John Locke. Locke has us imagine the fate of a Prince who one day wakes up in a strange bed dressed in ratty clothes. He discovers, to his shock and dismay, that the face he sees in the mirror and his surroundings are completely unfamiliar. Everyone he meets greets him as the town cobbler. He protests at great length that he is actually the Prince, but everyone takes him for a fool.

I take it that we can all conceive of the plight of the Prince, and we think that the prince is right in his assertions while the townsfolk are wrong in calling him a fool. In fact, this sort of scenario is a common theme in Hollywood films. Let's call scenarios of this sort **body switching cases**.

But notice that if our current definition is correct, these situations shouldn't make any sense. For the body that is causally continuous with the Prince's current body is that of the cobbler. So the person who wakes up in the cobbler's bed is the cobbler, not the Prince.

¹It is true that we will sometimes say things like, "He's a different person than he used to be" when someone undergoes a strong change in personality. But even this way of speaking seems to presuppose that there is *one* thing that has different personalities at different stages of its life.

²We don't mean to look for a definition of the word "person"; for that, we could go to any English dictionary. Instead, we're after what it is to *be* a person, the *essence* of personhood.

³This symbol stands for "if and only if". It means that whenever the condition to its left is true, so is the condition to its right, and vice versa. A and B stand for arbitrary individuals.

Examples such as these led Locke to presume that the definition of personhood has nothing to do with the body, but is instead something mental (or psychological).

1.2 The memory theory

Locke's positive proposal was that what sustains our identity through time is the presence of memories of our past selves. Thus, we can propose the following definition:

A is the same person as B iff A has memories of events in B 's life.

This theory can explain the body switching cases, as well as the fact I am the same person as the 14 year old who caught the fish in Alaska, because I remember doing so. But the definition is undermined by the simple fact (first pointed out by Thomas Reid) that **people can lose their memories** without losing their identity.

Reid imagines three stages in the life of a retired military general. At stage A , he is a young boy who steals from a neighbor's garden. At stage B , he is a middle aged military man who earns a medal of valor. And at stage C , he is an elderly man reflecting on his life. We are then asked to assume, which seems reasonable, that at stage B the general remembers stealing from the neighbor's garden, and at stage C , he remembers getting the medal of valor, but has forgotten his youthful indiscretion. If so, then Locke's theory gives the following results:

1. B remembers A , so $B = A$, and by symmetry, $A = B$.
2. C remembers B , so $C = B$, and by symmetry, $B = C$.
3. C doesn't remember A , so $C \neq A$, and $A \neq C$.
4. From 1, 2, and transitivity, $A = C$.
5. 3 and 4 are direct contradictions.

Thus, Locke's memory theory runs into a contradiction. This is enough to undermine the memory theory, but there is another interesting difficulty it faces. Imagine that A and B are best friends who spend every waking hour together. Later, after they have parted ways, A ponders their friendship (call this stage C). He fondly remembers everything they did together. The memory theory gives the following results:

1. C remembers A , so $C = A$, and by symmetry $A = C$.
2. C remembers B , so $C = B$.
3. A and B are distinct people, so $A \neq B$.
4. From 1, 2, and transitivity, $A = B$.
5. 3 and 4 are direct contradictions.

So, the memory theory also seems to run into trouble with **shared memories**. Now, this problem may seem silly because it's obvious that some of C 's memories are about *him* while others are about *his friend*. But saying what makes him *him* is exactly what the memory theory was supposed to do, so we can't presuppose such a notion in specifying which memories count without falling into circular reasoning.

1.3 The psychological theory

So, memories on their own won't cut it. But memories aren't our only mental states. And like we said with the body theory, it might not be the specific mental states we have that matters but that there is a certain continuity between them. So, maybe the following theory can finally get us what we need:

A is the same person as *B* iff *A*'s psychological state is causally continuous with *B*'s psychological state.⁴

This theory is probably pretty close to how most people think about the nature of personhood. But as we'll see the theory

2 Ego and bundle theories

If we take psychological continuity to give the identity conditions for persons, we are still left with a choice as to how to understand the notion of a person.

- The **ego** theory maintains that a person is the *subject* of experiences. In effect, the ego theory says that the reason continuity of psychological states makes for personal identity is that the person is an extra entity to whom the experiences belong.
- The **bundle** theory maintains that a person *just is* the set of continuous experiences. The reason continuity of psychological states makes for personal identity, is that there is nothing more to a person than a string of experiences.

2.1 Intuitions in favor of the ego theory

Teletransportation Imagine that you are offered the opportunity to travel to the Moon. The process is the following. You will step into a machine on Earth that will scan your entire body (brain included) to the tiniest detail, destroying it as it does so. The information recorded by the scanner will be radioed to the Moon, where a replicator will produce an exact duplicate of your body from materials held in a storage unit. Since the replica on the Moon has a brain exactly like yours, it will possess all your memories and will be psychologically continuous with you.

We then ask the question: Do you take the opportunity and step into the machine? Many people think that doing so is a sure way to die. It is only worthwhile to step into the machine if I will survive the process and wind up on the Moon. But surely I have no reason to care about the person who steps out of the replicator on the Moon. If you share this opinion, then you harbor ego theory tendencies. Since the person on the Moon is psychologically continuous with you, he isn't you only if there is some extra thing that hosts (is the subject of) those psychological states.

Slow replacement Imagine that over the course of some years you undergo a series of surgeries that replaces a portion of the cells in your body with exact duplicates. At the beginning of the process you retain most all your original cells. After a few surgeries, you have 50% new cells. And eventually all of your cells are replaced.

We then ask the question: After each surgery, who wakes up? It seems obvious that after the first surgery it is you that wakes up. But after the last surgery, the person that wakes up is just like the replica that steps out of the replicator on the Moon, so this wouldn't be you. So, at some point along the process the person that wakes up will no longer be you. If you share this opinion, then you harbor ego theory tendencies. Since the patient after the surgery is psychologically continuous with the patient before the surgery, the switch from *me* to *replica* can only take place if there is some extra entity that hosts those psychological states.

⁴We won't go into detail on exactly what it takes for one psychological state to be causally continuous with another. The answer will probably depend on how one thinks of the nature of mind and its relation to the brain.

2.2 Arguments for the bundle theory

Hume's empiricism We have no impressions (experiences) of the "self". We can only examine ourselves through our experiences; our selves are never laid bare to us, so to speak, absent any experience. In this way, the self (or the soul) is like the cause mediating between two conjoined events. We never see the cause, or the self, we only *infer* their existence. We take the self to be the thing that gives reference to our experiences; it is the subject of experience. But since we don't have experiences of the self directly, we cannot have knowledge of it.

And in fact, it does not seem that there could be an experience of the self. For the self is supposed to be some **continuous** and **invariable** entity that exists unified throughout our lives. But experiences are **fleeting** and perfectly **independent from each other**. No experience follows any other by necessity.

Thus, for all we have access to, we could be nothing more than a bundle of independent experiences following one after another. Much like he does for our ideas of cause and effect, Hume proposes a skeptical solution regarding our ideas of the self. From habit and the ease with which our mind transitions between experience, we posit a unifying entity underlying them.

Incoherence in the ego theory intuitions In the slow replacement case, we seem to think that there is a genuine question after each surgery who wakes up. After all, the person that wakes up can't be half-me, because I can't be divided up that way. But at what point does the switch occur. It seems that any point we pick is arbitrary; how could just a few more cells make the difference?

And we couldn't tell at what point the switch took place anyway, because the person who wakes up is psychologically continuous with the last person. So he will think that he is just waking up as himself.

The bundle theorist avoids these worries by denying that there is any question at all about *who* wakes up. Both answers are just different ways of describing the same situation. We know exactly what happens in the case without answering the question of who the person is. In a sense, there is no person beyond the chain of experiences had throughout the process of surgeries.

Split brain cases There are real life scenarios in which people have had the clump of cells joining the two halves of their brain (the *corpus callosum*) severed. Imagine that we show one of these people a different color in each half of their visual field, say red and blue. If we ask him to write the number of colors he sees with each hand, he will write "one" with each hand. But if we ask him to write the color that he sees, he will write "blue" with one hand and "red" with the other.

This suggests that there are two distinct **streams of consciousness** present within a single person. The psychological continuity theory says that a stream of consciousness is the mark of personal identity, but these cases seem to sever the connection between persons and streams of consciousness. If so, then what is the appellation "person" doing?

One could get around this worry by claiming that there are really two people in the one body of a split brain individual. But this response, too, faces difficulties.⁵

- Where was the second person before the corpus callosum was severed?
- How can a purely destructive act of cutting some brain cells bring a new person into existence?
- Which person now existing is the original person, and which is the brand new one?

The bundle theory avoids these issues by denying that there is any person there at all. There are simply two streams of consciousness, and that is all there is to the story.

⁵This isn't to say that these questions can't ultimately be answered. Perhaps you have an answer to them?

3 Consequences of the bundle theory

What is survival? The cases that pose trouble for the ego theory are far out there, but they have strong consequences for ordinary everyday events. The bundle theorist says that there is no real question whether you survive the teletransportation to the Moon. But what happens in that case is just what happens to us from one moment to the next. There is psychological continuity through any number of physical changes.

Thus the question of survival in the teletransporter case is the same as the question of survival from one moment to the next. And if we don't feel any strong connection to the person that emerges from the replicator, what connection ought we to feel to past and future stages that are psychologically continuous with us?

Does identity matter? The bundle theory says that the person before the severing of the corpus callosum is psychologically continuous with two streams of consciousness after the severing. But *identity* can never hold between one thing and two different things (it's always the case that if $A = B$ and $A = C$, then $B = C$). Thus, if what matters (what we care about) is psychological continuity, then identity isn't what matters. That is, it doesn't matter whether I am the *same thing* after some change.