

Lecture 2

Locke on Sortals and Psychological Continuity

1. Introduction

Locke is often credited as the first to give a psychological continuity account of personal identity. He is often described as giving a **memory criterion** for the persistence of persons over time. However, as we'll see, this may not be an accurate characterisation of his views in the *Essays*. Since the primary text is quite a challenging one to grapple with, we'll devote much of this lecture to a close reading of that text.

2. Locke on Identity (in General)

At the start of Chapter 27 (Book II) sets out what he takes to be the rules of that govern the notion of identity. He makes the following claim:

For we never finding, nor conceiving it possible, that two things of the same kind should exist in the same place at the same time, we rightly conclude, that, whatever exists anywhere at any time excludes all of the same kind, as is there itself alone. (§1)

He goes on to claim that there are three different kinds of substance: **God, finite intelligences** (or **souls**), and **bodies**. These three can exist in the same place at the same time. But there cannot be two of one *kind* of substance in the same place at the same time. For, he argues, if it is the case that two particles of matter can be in the same place at the same time, then it follows that two material bodies can be in the same place at the same time. But if this is possible, it "*takes away the distinction of identity and diversity of one or more and renders it ridiculous*" (§2).

Locke sets out two other rules for identity, or **principles of individuation** (i.e. *principium individuationis*). The three together are as follows:

- (i) No two things of the same kind can exist in the same place at the same time.
- (ii) No one thing can have more than one beginning
- (iii) No two things of the same kind can have a single beginning

Notice, though that these are claims *relative to a kind*. On Locke's view, individuation and identity conditions may vary depending on the relevant kind. In contemporary language, these kinds are sometimes referred to as **sortals**. So, if I want to know whether x is the same as y, I need first to answer the question "the same *what?*"

3. Distinguishing Kinds

Locke distinguishes between **human beings** (or “*man*”) and **persons**. He argues that when you ask about someone if they are the same thing as they were before, there are in fact **three** different things you could mean:

- (a) Is the **physical matter** the same collection of matter as it was before?
- (b) Is the **human being** the same human being as existed before?
- (c) Is the **person** the same person as they were before?

For a thing at t_1 to be the same **physical matter** (or “body”) as some thing at t_0 , it needs to be made of all and only the same atoms at t_1 as at t_0 . (§3)

For a thing at t_1 to be the same **human being** as some thing at t_0 , it needs to “participate in the same life” at t_1 as at t_0 . If you like, we can understand this in terms of continuity of biological function of the organism. The reason for this difference between humans and bodies is that, in general an organism can survive the loss of some parts of its body; but the body can never be exactly the same body when parts are lost.

Locke also argues against the view that x is a human being iff x is rational, since it is possible for there to be non-human animals that are rational, and it is possible for there to be non-rational human beings.

[...] whoever should see a creature of his own shape or make, though it had no more reason all its life than a cat or a parrot, would call him still a man; or whoever should hear a cat or a parrot discourse, reason, and philosophize, would all or think it nothing but a cat or a parrot; and say, the one was a dull irrational man, and the other a very intelligent rational parrot. (§8)

Locke then turns his attention to **persons**. He begins by stating “what person stands for”—i.e. what it is that personhood consists in.

*[person stands for] [...] a thinking intelligent being, that has reason and reflection, and can consider itself as itself, the same thinking thing, in different times and places; which it does **only by that consciousness which is inseparable from thinking**. (§9, my emphasis)*

In other words, x is a person iff x has consciousness. This is a claim about the **necessary and sufficient conditions** for being a person. But Locke then goes on to extend this into an account of **persistence** conditions.

For since consciousness always accompanies thinking, and it is that which makes everyone to be what he calls self, and thereby distinguishes himself from all other thing things, in this alone consists personal identity, i.e. sameness of a rational being: and as far as this consciousness can be extended backwards to any past action or thought, so far reaches the identity of that person (§9, my emphasis)

Thus, on Locke's view x is the same person as y iff x and y have the same consciousness.

Now, many have taken this to be equivalent to a **memory condition** of persistence. But it is important to note that for Locke, remembering does not exhaust the ways of having consciousness.

4. **Locke on Consciousness**

Following this statement of his position, Locke turns his attention to examining the conditions under which x and y have the same consciousness.

Forgetting (§10)

He immediately considers cases of forgetting. He acknowledges that there is "no moment of our lives wherein we have the whole train of all our past actions before our eyes in view". Here, he accepts that "our consciousness has been interrupted". So, what about things that we forget?

Locke considers a similar problem in §20, where considers a case where we lose memories irretrievably. Here again we ask, am I not the same person who performed those actions, despite forgetting them?

Yes, but this is not a problem, argues Locke.

He explains our great concern over this result by arguing that what we are doing is confusing the question "is x the same person as y?" with the question "is x the same thinking thing as y?", where by 'thinking thing' here Locke means to refer to the *soul*. (**NB:** Locke's use of the word 'substance' can be confusing to modern ears. He understands souls to be *immaterial substances*.) In cases of forgetting, we have the same soul, but are indeed different persons.

We may also, he suggests, be confusing 'I' as referring to the *person* who performed the action rather than as referring to the *human being* who performed the action. Our intuition is tracking sameness of human being, and then, "the same [human being] being presumed to be the same person" (§20) we take the forgetting cases to be instances of persistence of *persons* in the absence of consciousness. But this is mistaken.

Thus, on Locke's view **consciousness can be gappy**. For this reason, his view is not best understood as a *continuity* view at all. It is better understood as something like a "**Connected Consciousness Theory**" (from Kaufman (2016), "Locke's Theory of Identity").

Changes to physical body (§11)

Here Locke argues that the physical body can change while the same consciousness is preserved. What is particularly interesting about this section is that seems to include **proprioception** and other awareness of our body in the notion of consciousness. We have consciousness of our appendages and how they feel; nevertheless, if one of those appendages were severed from us, we would continue to be the same person.

Somewhat amusingly, later in §17 he writes:

Upon separation of this little finger, should this consciousness go along with the little finger, and leave the rest of the body, it is evident the little finger would be the person, the same person.

So, persistence of persons tracks the continuation of consciousness, wherever that consciousness goes.

'Body transplant' cases (§15)

Locke uses his case of the prince and the cobbler to demonstrate that you can have the same consciousness in two different human beings. For instance, if you were to swap the consciousnesses between two human beings (e.g. a prince and a cobbler), the person would go where the consciousness does, and so would be in a different human being. Thus, having the same body (the human being, rather than the particular physical particles at any given time) is not necessary for being the same person.

5. Body Transplants and Psychological Continuity

Locke isn't the only one to have taken body transplant cases to be evidence in favour of a psychological continuity view of persistence. Bernard Williams puts pressure on this view in his "The Self and the Future".

In that paper, he considers a few different variations on the transplant case:

First, suppose that the Prince and the Cobbler are told before the procedure that they will undergo a procedure wherein their psychological traits and memories will be transferred. Now suppose they are each told that the Prince's body will be tortured and the Cobbler's body will receive a large sum of money. Here, you'd expect the Cobbler, pre-procedure, to be very frightened, and the Prince (again pre-procedure) to be delighted. Moreover, after the

procedure, whoever spoke through the Price's mouth would say their fears were confirmed. And whoever spoke through the Cobbler's mouth would say they were right to look forward to their reward.

Now the second version. Forget everything you've just heard. If someone were to tell *you* your body will be tortured at some point in the future, it would be rational to fear it. Now suppose that someone told you that *before* your body is tortured, your memories will be wiped. That doesn't seem to be less reason to be fearful. Now, suppose they also tell you that, after having your memories wiped, but before your body is tortured, your body will receive a new set of memories. It still seems rational to fear the torture. Here then are all the things you're told in order:

- (a) Your body will be tortured
- (b) Your memories will be wiped before the torture
- (c) Your body will be implanted with new memories before the torture
- (d) The memories in (c) will be those that are currently in a living individual P
- (e) P will have your memories implanted in their body and will get a large sum of money afterward

In other words, it looks like we can just as well understand the mind-transplant case as a body-transplant case.

Where do we draw the line between survival and death in this case?