

10. Locke on Personal Identity

Martín Abreu Zavaleta

July 22, 2015

Locke is concerned with determining the conditions under which something persists through time. For the time being, we will understand this as the question under which conditions a given object at some point in time is *identical* to a given object at some future point. When Locke talks about the principle of individuation, he is talking about such conditions.

1 Preliminaries

Before considering the identity of persons, Locke considers the identity of other things: what he calls substances, plants, animals, and artifacts. He thinks that what makes substances persist through time is being composed of the same matter. Things are a bit more complicated when it comes to the identity of plants and animals.

According to Locke, a plant persists through time just in case it retains the functional organization of its components, even if it is not composed of exactly the same matter. He thinks the same is true of animals and some artifacts: “For Example, what is a Watch? ’Tis plain ’tis nothing but a fir Organization, or Construction of Parts, to a certain end, which, when a sufficient force is added to it, it is capable to attain.” (§5)

When it comes to a person, his whole discussion relies on distinguishing between levels of description. We may ask about the identity conditions for a person seen as a mere *substance*, whether it is the body or some thinking substance (what Locke would call a soul). At a higher level, we may ask for the identity conditions of a person, *seen as an animal*; this is what Locke calls ‘a man’. Finally, we may ask for the identity conditions of a person proper, with which Locke is mostly concerned in this chapter.

Before presenting Locke’s view, it will be useful to have some criteria for assessing it. We may call one such criterion *purely formal*. Locke has promised us a theory of what it makes a person at a given time identical to a person at some future time. He takes this task to be the one of offering a principle of individuation for people. If his theory succeeds, we should expect that his principle of individuation accounts for some formal properties of identity: reflexivity, symmetry, and transitivity.

A relation is *reflexive* just in case everything in its domain stands in the relation to itself. Identity is transitive because everything is identical to itself, but the relation of being someone’s parent is not, because no one is her own parent. A relation is *symmetric* just in case: if object X stands in the relation to object Y, then object Y stands in the relation to X. Identity is symmetric because if $X=Y$, then $Y=X$. The relation of being someone’s parent is not symmetric because if A is a parent of B, that doesn’t entail that B is also a parent of A. Finally, a relation is *transitive* just in case, if X stands in the relation to Y, and Y stands in the relation to Z, then X stands in the relation to Z. Identity is transitive, but the relation of being someone’s parent is not.

In principle, any good theory of personal identity should deliver the result that the relation *being the same person as* should be reflexive, symmetric, and transitive, if we are to take it seriously as an identity relation.

In addition to this formal criterion, we have a *substantive criterion*. This criterion demands that a theory of personal identity deliver the right predictions or judgments on whether two given people at different times are one and the same person. We may understand the substantive criterion as the requirement that the theory of personal identity under consideration is (a) not subject to counterexamples, and (b) comprehensive enough to cover all the cases in which two people at different times are one and the same.

2 Locke's views on personal identity

Here is a passage from Locke's book:

[S]ince consciousness always accompanies thinking, and 'tis that, that makes every one to be, what he calls *self*; and thereby distinguishes himself from all other thinking things, in this alone consists *personal identity*, i.e. the sameness of 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*; it is the same *self* now it was then; and 'tis by the same self with this present one that now reflects on it, that that Action was done. (§9)

What Locke says is that a person later in time will be the same as some person earlier in time just in case the latter remembers everything that happened to the person earlier in time. Let's make this more precise by saying that, according to Locke, a person A at time t_0 is the same as person B at time t_{0+n} just in case, at time t_{0+n} , person B remembers all the experiences and thoughts that A went through at t_0 . Locke's criterion of identity is sometimes called *the psychological criterion of identity* for obvious reasons.

For the rest of the chapter, Locke is concerned with making some precisions and presenting some of the consequences of his view. We'll see if these consequences are acceptable from the point of view of the substantive and formal criteria, respectively:

As Locke is quick to admit, his view on personal identity entails that someone who is asleep and is not dreaming is not the same person as the one who went to bed that night, or the one that will wake up in the morning. The person who goes to bed and the person who wakes up might well be one and the same, but the person who is asleep is different. Moreover, Locke seems to say that, strictly speaking, there is no person who is asleep, since to be a person is to be conscious and having thoughts, perceptions, and the like. On the assumption that dreamless sleep is devoid of thoughts, perceptions, and the like, there is no consciousness present when a body sleeps.

Ordinarily, we wouldn't think that someone ceases to exist merely because she went to sleep. In fact, we often speak as if the person who is asleep is one and the same with the one who went to bed and the one that will wake up in the morning. We say things like "I didn't hear the phone because I was asleep". If Locke is right, those things are, strictly speaking, false, since a person is not present during sleep. As far as the substantive criterion goes, Locke's view doesn't seem to be doing well.

Locke is also very clear that, on his view, it is possible for a single person to inhabit two different bodies at two different times. This would happen if somehow we managed to transfer all the

memories from one body to another, without leaving any trace in the original body. Similarly, if immaterial souls, rather than bodies, are the places where we store our memories and that have our sensations, if we managed to transfer someone's memories from one soul to another, Locke thinks that the same person would have inhabited one soul originally, but another after the transfer.

Perhaps it is because Locke's ideas have laid roots into our culture, but it's difficult to resist this sort of conclusion. Sometimes with comical intentions, some fictional works present us with a situation in which two people swap bodies. In those kind of situations, the person with one body remembers many of the events that happened to the other body from a first-person perspective, and have the inclinations and character that the other body used to have, etc. So insofar as our intuitions concerning personal identity go, this consequence of Locke's view seems unproblematic.

Forgetfulness and amnesia seem to present more pressing challenges. Locke's criterion of identity requires a person to have the same memories as another one. However, I may easily forget certain details of what happened yesterday, or even what I had for breakfast this morning. If we adopt Locke's demanding criterion of identity, we will be forced to conclude that I am not the same person as the one that had breakfast this morning or that was teaching class yesterday. But again, this seems contrary to some very natural intuitions.

A more radical case is one in which someone has amnesia. If Locke is right, if I have an accident and as a result I get amnesia, I won't be the same person as the one that had the accident. This is not merely a metaphorical way of speaking: Locke quite literally says that I wouldn't be the same person. **Question:** What would happen if all of a sudden I recovered my memory? Who would I be?

Whatever one thinks of the amnesia case, surely I won't be a new person just because I forgot what I had for breakfast, so perhaps we should amend Locke's criterion. Perhaps we shouldn't demand that a person remembers *everything* that happened to her past self, but only that she remembers *enough* of the things that happened to her past self.

Unfortunately, this still seems to be too restricting. In a single day, I have many different experiences of which I can only remember very few, so even this new criterion won't deliver the right results. One way of amending this criterion is to say that identity between people at different times merely requires remembering some important events, or some set of events suitably defined. As long as we have some independently motivated way of determining those important events, it seems we can save the spirit of Locke's view from counterexamples.

Later in the class we'll see some more sophisticated counterexamples to Locke's thesis, but for now let's see if it can satisfy the formal requirements, starting with reflexivity. Because identity is reflexive, everyone is identical to herself. Is it true that, at any given time, a person has exactly the same memories as herself at that same time? Surely it is, so Locke's view at least respects the reflexivity of identity.

Things are not so easy with symmetry. Consider John at time t_0 and John at time t_{10} , and suppose that John at t_{10} remembers everything that happened to John at t_0 . Unfortunately for Locke, this doesn't entail that John at t_0 remembers everything that happened to John at t_{10} , since those things hadn't happened yet! So the relation of remembering everything that happened to someone is not symmetric. **Question:** How do you think that Locke could solve this problem?

Locke's theory doesn't have a better Locke with transitivity. The British philosopher named Thomas Reid noticed this and wrote about it in one of his *Essays on the intellectual powers of men*. He presents us with the following case:

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. When 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 as him who was flogged at school (Essays, 276).

Reid argues against Locke by *reductio ad absurdum*: he shows that from Locke's theory together with some plausible premises—in particular, the transitivity of identity—we can derive a contradiction. Since there are no true contradictions, we must reject one of the premises. We can't reject the transitivity of identity, so instead we must reject Locke's theory.

What Reid's example illustrates is simply that the relation of remembering everything that happened to someone is not transitive. In this case, the general remembers what happened to the official, and the official remembers what happened to the boy, but the general doesn't remember what happened to the boy. Because the relation of remembering what happened to someone else is not transitive, Locke's view also fails in this respect.

There are ways out of these objections. What we need is to find a relation that is close enough in spirit to the relation Locke offers, but that unlike that relation, is both transitive and relation. Let's start with transitivity: traditionally, philosophers observed that even though the relation of remembering what happened to someone is not transitive, there is another relation in the vicinity that is transitive: let's call that relation *psychological continuity*, even though other people may not use the expression that way.

In Reid's case, the general remembers what happened to the official and the official remembers what happened to the boy. So the general stands in the following relation to the boy: the relation of remembering what happened to someone who remembers what happened to him (the boy). If this is the relation that defines personal identity, we don't need to worry about Reid's case anymore. Unfortunately, our new relation can only be a provisional solution: if we have a long enough series of people that is relevantly like the series presented by Reid, that will be enough to raise a new objection.

Luckily, we can generalize this newly introduced relation to avoid those problems. Say that A is *psychologically connected* to B just in case A remembers everything that happened to B—we could relax the relation and just ask that A remembers most of what happened to B, or the important things that happened to B, etc. This is basically our old, non-transitive relation, but we can use it to define a new, *transitive* relation. Say that A is *psychologically continuous* with B just in case they belong to a series of people each of which is *psychologically connected* to her predecessor. This new relation is transitive.

Notice that our newly minted relation also solves the problem with symmetry. For if A belongs to the same series of psychologically connected people as B, then B belongs to the same series of psychologically connected people as A. So perhaps Locke's theory was wrong in the letter, but at

least so far we have managed to deflect some of the objections against its spirit coming from the formal requirements.

3 Further objections

Perhaps the criteria above have been successfully met, but that doesn't mean there are no more objections. One famous objection is known as *the circularity objection*. This new objection is not concerned with the truth of Locke's theory. Instead, it is concerned with whether it is explanatory or informative.

By the beginning of the course, we saw some cases of definitions that were not explanatory. For instance, it's not explanatory or informative to define chairs as the kinds of things that fulfill the function of chairs. The circularity objection is of this kind. It points out that the relation of remembering something *presupposes* identity: you simply can't remember what happened to other people. You can remember seeing what happened to other people, of course, but that's not the point. You can't remember the experiences of other people, or doing things that *you* didn't do. This is usually taken to be a fact about what it is to remember something.

But if this is true, then Locke's account is not explanatory. If it succeeds, it only succeeds because it uses a relation that itself presupposes identity, in which case Locke hasn't really explained what grounds that identity itself.

Some people have tried to amend Locke's view by changing the kinds of mental states that are required for psychological continuity. Derek Parfit is among the people who discuss this new view. The new view requires the introduction of a new word, 'quasi-memory'. A person *quasi-remembers* something just in case she has something that is qualitatively indistinguishable from a memory, but that doesn't presuppose identity. For instance, I could quasi-remember everything that happened to you if someone makes some sort of copy of your brain and then puts that information into my brain. Parfit then defines psychological continuity in terms of *quasi-memories* rather than in terms of genuine memories.

This new theory apparently can solve this version of the circularity objection, but it seems to have deeper problems. For surely being psychologically continuous in this new sense is not sufficient for identity. If someone leaves my memories intact and just inputs some quasi-memories of yours into my brain, that doesn't mean that I'm identical to you. I may, for instance, be perfectly aware that these are not genuine memories, or be perfectly aware that I never did the things that the quasi-memories represent their holder as doing, yet still have those quasi-memories. We'll continue to discuss some interesting issues concerning personal identity in future sessions.