Class 13 - March 2
Locke's Theory of the Self

I. Body and Soul

We have discussed two accounts of personal identity: the body theory and the soul theory.
According to the body theory, we are our bodies.
The body theory is undermined by the inconstancy of material constitution.
Our bodies change quickly, and thoroughly.
We are all like Gregor Samsa, physically, in the sense that our material constitution changes constantly, even if our appearance does not.
The changes to our selves are more like the psychological or experiential changes to Gregor, gradual and almost imperceptible.
Moreover, we might believe that there is a constant, underlying self, a haecceity which has the experiences, and undergoes the changes.
In that latter case, the body theory has no plausibility.

The soul theory, at which Descartes hinted, and which we saw explicitly in Plato's *Phaedo*, claims that we are essentially thinking things, our souls.
I mentioned that Plato ascribed different properties to the soul than Descartes did.
But, according to any version of the soul theory, the self is an immaterial substance, a thing completely distinct from our bodies.
There are two sorts of objections to the soul theory.
The first sort of objection denies that there are souls distinct from bodies.
This (usually-materialist) response argues that the soul theory must be false since there are no souls.

One argument for the first kind of response relies on a problem of interaction between the body and the soul.
We will discuss this problem in more depth after break, when we look at the philosophy of mind.
Briefly, the problem is that if the soul is immaterial, it seems impossible for it to interact with the body in the way that our minds and our bodies do relate.
For Plato, the problem of interaction is not a serious problem, since he denies that the world of bodies is a real world.
I have been trying to put aside questions of skepticism about the material world.
Descartes may be interpreted as trying to solve the problem of interaction by claiming that the self is an amalgam of mind and body.
He even locates the soul in the pineal gland.
But, that just pushes the problem back, since a soul, in his sense, is non-physical and so can not be located anywhere.

The second sort of objection to the soul theory of self, which we see in Locke's work, does not oppose the existence of souls.
Locke points out that the defender of the soul theory is committed to the independence of bodies and souls.
This, the body in which the soul is placed is inconsequential.
The same soul could be put into two bodies.

> If the identity of *soul alone* makes the same *man*, and there be nothing in the nature of matter why the same individual spirit may not be united to different bodies, it will be possible that those men, living in distant ages, and of different tempers, may have been the same man: which way of speaking must be from a very strange use of the word man, applied to an idea out of which body and shape are exclude[d]... (Locke 340)

Imagine that a soul had two different incarnations.
We wouldn't say that there were only one person.
Here is Locke, in a section you did not read, providing an example.

> Suppose it to be the same soul that was in *Nestor* or *Thersites* at the siege of *Troy*...which it may have been, as well as it is now the soul of any other man. But he now having no consciousness of any of the actions of either of *Nestor* or *Thersites*, does or can he conceive himself the same person with either of them? Can he be concerned in either of their actions, attribute them to himself, or think them his own more than the actions of any other men that ever existed? Thus, this consciousness not reaching to any of the actions of either of those men, he is no more one self with either of them than if the soul or immaterial spirit that now informs him had been created and began to exist, when it began to inform his present body... (Locke, *Essay* II.XXVII.14).

The soul is not identical to the self, Locke says, since there can be more than one self using the same soul.
Moreover, Locke argues that the problem works in the other direction, too.
Locke believes that it is possible for the same self to be transferred between souls, as we will see when we get to Locke's positive account of the self.

So, we can distinguish between two different types of objections to the soul theory, one which accepts the existence of souls and one which rejects them.
On either objection, the soul theory of self meets counter-intuitive consequences.
To avoid these unfortunate consequences, Locke provides a different account of self.
Locke's account is both controversial and revolutionary.

II. Identity and Sortals

Locke's account of personal identity is important for two reasons.
Neither of those reasons is that it is clear and well-written.
Locke's work is turgid and often nearly incomprehensible.
To be fair, Locke himself recognized the problem.
Still, it doesn't make his work any easier to read.

One of Locke's lasting contributions to the literature on personal identity is his observation that identity generally is relative to a sortal, to a kind of thing.

> We must consider what is meant by Socrates, or the same individual *man*. First, it must be either the same individual, immaterial, thinking substance; in short the same numerical soul, and nothing else. Secondly, or the same animal, without any regard to an immaterial soul. Thirdly, or the same immaterial spirit united to the same animal (Locke 343).

Let's say that my daughter re-forms her plasticine sculpture of a horse into the shape of a house.

The lump of plasticine is the same lump, but it is a different statue.
The ship of Theseus may be the same ship while being a different material object.
We can not know how to identify something unless we know what kind of thing it is.
The same mass of matter may be a different statue while being the same toy.
So, we can not know what our identity is until we know what kind of thing we are.

We might, for example, think that we are a biological kind of thing.
Locke takes 'man' to refer to a type of animal.

> The idea in our minds, of which the sound "man" in our mouths is the sign, is nothing else but of an animal of such a certain form (Locke 341a).

An animal is not merely its matter.
The matter remains after the animal's death while the animal itself does not.
The principles of identity of the sort 'man' are biological.
The identity of a man is determined functionally, by its organization and not by its matter.
Note that a body theorist of the self could make this kind of appeal to a biological sortal.
The body theorist could say that we are men, material objects with a certain sort of functional organization.

But, according to Locke, the sort man can not serve as the sort of our selves.
A man is identified by the functional organization of the body; it is a biological thing.
But, 'person' and 'self' are forensic, or moral, terms used for practical purposes of ascribing responsibility.
One can see an argument for the distinction between (biological) humans and (moral) persons clearly when we consider the question of whether aliens, or sentient machines, could be persons.
Since such a case is possible, our personhood must not be identical with our biology.

In addition, Locke's account arises in part from the worry, raised by Robert Boyle, about the resurrection of bodies in the presence of cannibals.
Imagine that some portion of one person's body is eaten by another person, and so becomes part of both of them.
It is a puzzle to determine into whose body that portion of matter will go at the resurrection.
I'm not sure how much Locke's rejection of the body theory and his claim that 'man' is the proper sortal for personal identity are responses to Boyle's worry.
But, Boyle's puzzle seems to have had some influence on Locke.

We are looking for a criterion for identity for personhood: what makes us the same people over time.
It's not sameness of body.
It's not sameness of soul.

III. The Consciousness Theory

For Locke, what makes the same person over time, is consciousness, and, especially, connection through memory, which Locke calls consciousness extending backwards.
Locke's view is sometimes labeled the psychological continuity theory, and sometimes the memory theory.
I'll call it the consciousness theory.

> [A person] is 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, and, as it seems to me, essential to it... (Locke 341a).

Locke argues for the consciousness theory from premises similar to those that Descartes invoked for the soul theory.
In thinking about ourselves, we think about our thoughts.
For Descartes, consciousness is the essential characteristic of mental life, and what distinguishes us from (other) animals.
Locke denies Descartes's conclusion that we are our souls.
But, he maintains an emphasis on conscious thought.

> Since consciousness always accompanies thinking, and it is that which makes every one to be what he calls self, and thereby distinguishes himself from all other thinking things, in this alone consists personal identity (Locke 341a).

Locke's view is called the psychological-continuity theory because of its claim that continued consciousness is a mark of sameness of self.
We know of continued consciousness through memory.
So, memory is also essential to the criteria for identifying ourselves over time.

> For as far as any intelligent being *can* repeat the idea of any past action with the same consciousness it had of it at first, and with the same consciousness it has of any present action; so far it is the same personal self (Locke 341b).

Locke's argument for the consciousness theory also invokes a series of thought experiments.
He considers a prince who transfers his consciousness to a cobbler.
Though he inhabits a different human being, Locke argues, it is the same prince in the pauper's body.
Reid disagrees.

> This doctrine has some strange consequences...Such as, that if the same consciousness can be transferred from one intelligent being to another, which he thinks we cannot show to be impossible, *then two or twenty intelligent beings may be the same person* (Reid 347a).

Reid and Locke are here trading intuitions.
My intuitions favor Locke's account.
But, they shouldn't be taken as decisive.
Ask yourself what you think about the case.

Elsewhere, Locke considers a case of a day and night man.
Consider a single man who has one consciousness in the day and one in the night, like Jekyll and Hyde.
In this case, Locke says, we are tempted to say that there are two people in one biological man.

Locke's solution to the problem of personal identity helps explain his objection to the soul theory.
The soul theory posits that sameness of soul, taken as a substance, suffices for sameness of person.
Locke considers a case in which consciousness varies but substance remains.
If the soul theory were correct, then we should have the same person.
But, Locke says that we have two different people, and so the soul theory is wrong.

> If the same consciousness...can be transferred from one thinking substance to another, it will be possible that two thinking substances may make but one person. For the same consciousness being preserved, whether in the same or different substances, the personal identity is preserved (Locke, *Essay* II.XXVII.13).

IV. Reid Against Locke

Locke's account of the self relies on continued consciousness.
We know that we are the same person who had a certain earlier experience when we can recall, using memory, that experience.
Memory is essential to connecting our consciousness.

Reid worries about Locke's blurring the difference between memory and consciousness.
Locke discusses memory as if it is consciousness of a past experience.
Reid notes that there is a difference between remembering an experience and having consciousness of that experience.

> When, therefore, Mr. Locke's notion of personal identity is properly expressed, it is, that personal identity consists in distinct remembrance; for, even in the popular sense, to say that I am conscious of a past action means nothing else than that I distinctly remember that I did it (Reid 347b).

Even Reid notes that this objection is not a serious one; it's just an unfortunate use of language which seems to have no effect on the theory of self.
The more serious concern, as Locke himself points out, is that there are gaps in both our conscious experience and in our memory.
Every time we sleep, we lose consciousness.
Some experiences are forgotten.
Two things can only be identical if they have a continued existence.

> While any being continues to exist, it is the same being; but two beings which have a different beginning or a different ending of their existence cannot possibly be the same (Reid 346a).

If consciousness and psychological continuity are required for personal identity, then every time we sleep or lose a memory, we lose our identity.

> If the intelligent being may lose the consciousness of the actions done by him, which surely is possible, then he is not the person that did those actions; so that *one intelligent being may be two or twenty different persons*, if he shall so often lose the consciousness of his former actions (Reid 347a).

This worry is perhaps better expressed in the problem of the old general.
The old general remembers being a middle-aged officer.
The middle-aged officer remembers an experience from his childhood, being flogged for robbing an orchard.
But, the old general does not remember being flogged.
According to the transitive property of identity, the old general is surely the same person as the child.
But according to Locke's theory, it seems, the old general is not the same person as the child.

The old general cannot remember being flogged as a child.

> 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 (Reid 347a).

Reid has thus derived a contradiction from Locke's theory.
The problem expands, since our consciousness is not a constant, but a stream of changing experiences.

> Our consciousness, our memory, and every operation of the mind, are still flowing like the water of a river, or like time itself (Reid 348a).

Since our conscious experiences change constantly, it looks like Locke's theory entails that we are constantly changing, too.

Reid's worry about these oddities may not be decisive against Locke.
They point out some counterintuitive consequences of the consciousness theory, and in the case of the contradiction force Locke to restate some of his claims.
But perhaps our conscious experience, ever flowing, does not support any kind of sameness of an individual over time.
Reid suggests that Locke assert that we are the same kind of individual, rather than the same individual, over time.

Even Reid's strongest words about these related problems do not show that he took them as refuting Locke.

> Is it not strange that the sameness or identity of a person should consist in a thing which is continually changing, and is not any two minutes the same (348a)?

Most philosophical views have some strange consequences.
The question is whether their virtues outweigh their oddities.

V. Reid on Evidence

Lastly, Reid has a more-significant complaint about Locke's theory.
He claims that Locke confuses personal identity with evidence for personal identity.
My memories of an experience are evidence that I had that experience.
My consciousness of an experience is evidence that I am having an experience.
But, my identity is not my memory or my consciousness.

> It is to attribute to memory or consciousness a strange magical power of producing its object, though that object must have existed before the memory or consciousness which produced it (Reid, 347b)

Reid illustrates the formal problem, which is a general one, using the example of a stolen horse.
The owner of a stolen horse, when presented with a horse which might be the stolen one, will use its similarity to the original to determine if the two horses are the same.
The similarity of the original horse and the present one is evidence that the two horses are the same.

> But would it not be ridiculous from this to infer that the identity of a horse consists in similitude only (Reid 348a)?

To put Reid's point in some philosophical jargon, he is accusing Locke of arguing from epistemological premises to metaphysical conclusions.
We discover, says Reid, that we are the same over time by using memory and consciousness.
But we should not confuse the way we learn about something with the thing itself.
We use memory and consciousness to identify our selves, but our selves are not merely continued consciousness or memory.

A more complete discussion of the lessons that Reid takes from the considerations of evidence will have to wait until next class, when we look at Reid's positive account of personal identity, which we might call a non-reductive theory.