Class 14 - Identity and the Self
Locke, *An Essay Concerning Human Understanding*
Book II, Chapter XXVII (AW 367-377)

I. The Mind-Body Problem

While Locke was suspected of Hobbesian materialism, he is actually a dualist, accepting the existence of the soul and God, as well as material objects.
So Locke, like Descartes, is saddled with a mind-body problem.
The slight metaphysical differences between Locke and Descartes do nothing to change the essential nature of the problem.
Given what Locke says about the primary qualities, we can see that bodies have several essential properties, not merely extension.
Similarly, minds are not essentially thinking; they are just the kinds of things that do think.
Still, nothing in these changes in the characterizations of each substance mitigates the problem.

Locke does not provide a Cartesian-style solution to the mind-body problem, despairing of any satisfactory account.
His discussion of mind-body interaction is humble, though, rather than skeptical.

> Supposing the sensation or idea we name whiteness be produced in us by a certain number of globules, which, having a verticity about their own centres, strike upon the retina of the eye, with a certain degree of rotation, as well as progressive swiftness; it will hence easily follow, that the more the superficial parts of any body are so ordered as to reflect the greater number of globules of light, and to give them the proper rotation, which is fit to produce this sensation of white in us, the more white will that body appear, that from an equal space sends to the retina the greater number of such corpuscles, with that peculiar sort of motion... I cannot (and I would be glad any one would make intelligible that he did), conceive how bodies without us can any ways affect our senses, but by the immediate contact of the sensible bodies themselves, as in tasting and feeling, or the impulse of some sensible particles coming from them, as in seeing, hearing, and smelling; by the different impulse of which parts, caused by their different size, figure, and motion, the variety of sensations is produced in us (IV.II.11, not in AW).

Locke describes how the communication of motion of light to the retina will impel us to see a color, without any serious metaphysical speculation about the communication of motion.
He provides no explanation of why a particular wavelength of light should correspond to a particular color, or why the vibration of the ear at should correspond to the sound it does.
Why do lemons appear yellow?
Both the moderns like Locke and contemporary neuroscientists lack an explanation of the connection between my conscious experience and its cause.
Why is it that such and such motions in the air cause me to hear a symphony?
Why is it that certain wavelengths of light cause me to see blue?

> That the size, figure, and motion of one body should cause a change in the size, figure, and motion of another body is not beyond our conception. The separation of the parts of one body upon the intrusion of another and the change from rest to motion upon impulse, these and the like seem to have some *connection* one with another. And if we knew these primary qualities of bodies, we might have reason to hope we might be able to know a great deal more of these

> operations of them one upon another. But our minds not being able to discover any *connection* between these primary qualities of bodies and the sensations that are produced in us by them, we can never be able to establish certain and undoubted rules of the consequence or *coexistence* of any secondary qualities, though we could discover the size, figure, or motion of those invisible parts which immediately produce them. We are so far from knowing what figure, size, or motion of parts produce a yellow color, a sweet taste, ro a sharp sound that we can by no means conceive how any *size, figure*, or *motion* of any particles can possibly produce in us the *idea* of any *color*, *taste*, or *sound* whatsoever; there is no conceivable *connection* between the one and the other (IV.III.13, AW 395b-396a).

This is the kind of statement that worries those of us who like to think of philosophy as a science-like discipline, seeking progress on specific problems.
We haven't made any progress in the last three centuries of trying to answer that question.
That question is essentially what David Chalmers calls the hard problem of consciousness.
The easy problem is to map the brain, and to know all its functions.
Once we have done that, though, we still won't be any closer to an answer to why certain neural firings correspond to certain conscious experiences.

Locke's claim concerning our conscious experience is merely that there are lawful correspondences between physical events and some mental states.
If these lawful correspondences are possible, it seems possible for matter to think.
Moreover, it seems equally unlikely for whatever substance in which thought resides to be the seat of thought as for matter to be the seat of thought.

> We have the *ideas* of *matter* and *thinking*, but possibly shall never be able to know whether any mere material being thinks or not, it being impossible for us, by the contemplation of our own *ideas*, without revelation, to discover whether omnipotence has not given to some systems of matter fitly disposed a power to perceive and think, or else joined and fixed to matter so disposed a thinking immaterial substance - it being in respect of our notions not much more remote from our comprehension to conceive that God can, if he pleases, superadd to matter a faculty of thinking than that he should superadd to it another substance with a faculty of thinking, since we do not know in what thinking consists, nor to what sort of substances the Almighty has been pleased to give that power... (IV.III.6, AW 393b).

Locke thus draws a humble conclusion.

> *The extent of our knowledge* comes not only short of the reality of things, but even of the extent of our own *ideas* (IV.III.6, AW 393a).

The materialist and the dualist each make the error of claiming to know something that is beyond the reach of our ideas.

II. Personal Identity

The question of how to define or characterize personal identity, what identifies us or makes us the same over time, is deep and compelling.
Two standard, if limited accounts, might be called the body theory and the soul theory.

According to the body theory, which we might attribute to Hobbes, we are our bodies.
This view would be consistent with our general, contemporary preference for materialism.
The problem with the body theory is that our bodies are changing all the time.
We lose skin and hair constantly.
Every seven years, all the cells in our bodies are replaced.
So, if we identify ourselves with our bodies, we are not the same person we were, say, a moment ago.
(And, my son is made out of chicken nuggets and noodles.)

The problem of personal identity is related to a more general problem called the problem of material constitution.
Consider the ship of Theseus.
We can replace every plank on the ship, one at a time.
It could change its material composition completely, but remain the same ship.
We can even make a new ship with the old wood and find ourselves confused about what to say.
Is the ship that Theseus uses, with all new materials, his ship?
Or, is the new ship made of the old wood his ship?

The body theory is undermined by the inconstancy of material constitution.
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 can ascribe to Leibniz as well, claims that we are essentially thinking things, our souls.
According to the soul theory, the self is an immaterial substance 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 the problem of interaction.
If we identify ourselves with our souls, then we have a puzzle about how we interact with our bodies.

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.

> Souls [are], as far as we know anything of them, in their nature, indifferent to any parcel of matter... (§II.XXVII.14, AW 372a).

Thus, 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. This 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 excluded (II.XXVII.6, AW 369a).

Imagine that a soul had two different incarnations.

We wouldn't say that there were only one person.

> 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... (II.XXVII.14, AW 372a).

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, which we can call the consciousness theory.
Locke's account is both controversial and revolutionary.

III. Identity and Sortals

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 can not know what our identity is until we know what kind of thing we are.
There are three main types of things for Locke, as for Descartes: God, finite minds (souls), and bodies.
These categories are too coarse to help us identify our selves.

> 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 (II.XXVII.21, AW 374a).

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, like 'human being'.

> 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 (II.XXVII.8, AW 369b).

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.
Biological criteria are not strictly material.
The identity of a man is determined functionally, by its organization and not by its matter.

> The identity of the same *man* consists...in nothing but a participation of the same continued life, by constantly fleeting particles of matter, in succession vitally united to the same organized body (II.XXVII.6, AW 369a).

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, when souls are supposed to be reunited with their bodies.
Does it go with the eaten person or with the cannibal?
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.

IV. 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 called 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... (II.XXVII.9, AW 370a).

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 (II.XXVII.9, AW 370a).

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 (II.XXVII.10, AW 370b).

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.

Locke also considers the case of a day and night man, 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.
He is taking the soul to be a conduit for thought and imagining that different consciousnesses (especially memories) are placed in the same container soul.
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 (II.XXVII.13, AW 371b).

Locke's consciousness theory leads to some counter-intuitive consequences.
There are gaps in both our conscious experience and in our memory.
Every time we sleep, we lose consciousness.
Some experiences are forgotten.
If consciousness and psychological continuity are required for personal identity, then every time we sleep or lose a memory, we lose our identity.

Thomas Reid expresses this worry in a case called the Brave Soldier.
An old general remembers being a brave officer.
The brave soldier 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.
Reid's concern expands, since our consciousness is not a constant, but a stream of changing experiences.
Since our conscious experiences change constantly, it looks like Locke's theory entails that we are constantly changing, too.

Reid's worries point out some counterintuitive consequences of the consciousness theory.
Perhaps our conscious experience, ever flowing, does not support any kind of sameness of an individual over time.

The soul theory of self posits a constant haecceity, a thing underlying our experiences.
Locke's consciousness theory attempts to maintain a consistent thing, the self, without positing a particular substance called the self.
Another option, which we will see in Hume's work, is to give up the notion of a self in response to the constant changes of both material constitution and conscious experience.