Class 15 - The Bundle Theory
Hume, "The Self"
Parfit, "Divided Minds and the Nature of Persons"

I. Five Failed Theories of the Self

We have considered five distinct theories of the self: the body theory, the biological theory, the soul theory, the memory theory, and the irreducibility theory.
On the implausible body theory, we are our material constitution.
That constitution is constantly changing in a way that our selves are not changing.

On the more-plausible biological theory, which Locke called the identity of man, we are our biological selves: the body considered according to its functional organization.
The biological theory of personhood applies the same identity conditions to people as we do to other individual animals.
Locke provides thought experiments, including the Prince and the Cobbler, which ask us to consider the possibility of transferring consciousness among biological entities.
Most of our intuitions support the claim that the self transfers with consciousness and memory.
The person consisting of the Cobbler's body and the Prince's thoughts is the Prince and not the Cobbler.

Similar arguments, from Locke, apply to the soul theory.
Given some common views about the soul, that it exists prior to birth and after death, our personhood can be seen to be different from our soul by considering some simple thought experiments.
We can imagine two different souls inhabiting (or whatever the relation is) the same person.
Two different persons can house (or whatever) the same soul.

Locke's consciousness, or memory, theory relies on psychological continuity to identify the person over time.
Reid showed that problems with memory and gaps in consciousness entail that, on Locke's theory, we are disjoint.
The old general is not the same as the young child, if his memory is discontinuous.
We tried to save Locke's theory by appealing to psychological continuity, perhaps identified with some neurological account of subconscious thought and memory.
But further worries about the construction of memory can undermine even the fortified Lockean theory.

Instead of Locke's theory, Reid defends an irreducibility theory of self: the self is what it is, and can not be reduced to any other property or substance.
The irreducibility theory accepts the existence of a self, but denies that we can do anything but characterize our selves, in various ways.
The lack of a reduction makes the irreducibility theory awkward.
It gives up explaining the identity of the self while at the same time asserting that we have personal identity.
Kripke gave us hints that we have some essential characteristics, but not a full-fledged theory of the self.

Another option, for those who deny the reducibility of the self, is to wonder if there is a self at all.
We never sense our selves.
We sense our bodies, but both the body theory and the biological theory seem insufficient as accounts of our selves.

Berkeley, whose work on the self we are not reading, argues that since all ideas must come from sensation, we have no idea of the self, but that doesn't stop him from believing that there is a self. Berkeley, like Plato and Descartes, identifies the self with the soul.

> There can be no idea formed of a soul or spirit; for all ideas whatever, being passive and inert... they cannot represent unto us, by way of image or likeness, that which acts...The words *will*, *soul*, *spirit* do not stand for different ideas or, in truth, for any idea at all, but for something which is very different from ideas, and which, being an agent, cannot be like or represented by any idea whatsoever - though it must be admitted at the same time that we have some notion of soul, spirit, and the operations of the mind, such as willing, loving, hating, inasmuch as we know or understand the meaning of those words (Berkeley, *Principles* §27).

Hume and Berkeley, despite very different views of the world, agree on epistemological principles. Hume's view of the self, which follows Berkeley's claims and arguments, is recalled in Parfit's view as well.
Where Berkeley claims that there is a self even though we lack an idea of it, Hume and Parfit deny that there is a self in any serious sense.

II. Hume's Bundle Theory

Like Locke and Hume, Berkeley thought of ideas as pictures in the mind.
He concludes from the observation that we have no idea of the self that there are things of which we have no ideas, no pictures in our minds.
For Berkeley, we have no idea of God and no idea of the self, but we still have some notion of both, and no good reason to deny the existence of either one.

Hume agrees with Berkeley that we have no sense impression, and thus no idea, of the self.

> It must be some one impression, that gives rise to every real idea. But self or person is not any one impression, but that to which our several impressions and ideas are supposed to have a reference. If any impression gives rise to the idea of self, that impression must continue invariably the same through the whole course of our lives, since self is supposed to exist after that manner. But there is no impression constant and invariable. Pain and pleasure, grief and joy, passions and sensations succeed each other and never all exist at the same time. It cannot, therefore, be from any of these impressions or from any other that the idea of self is derived, and, consequently, there is no such idea (Hume 349a).

Unlike Berkeley, Hume concludes that we have no reason to believe in the self.
In other words, Berkeley argues B where Hume defends H.

B
B1. We have no idea of the self.
B2. But there is a self.
BC. So, there are things of which we have no idea.

H
H1. If we can have no idea of something, then that thing does not exist.
H2. We have no idea of the self.
HC. There is no self.

If what we mean by the self is some constant substance or property which persists through time, there is no such thing.
There is no underlying, unifying object which we can call the self.
There are just perceptions.

> When I enter most intimately into what I call *myself*, I always stumble on some particular perception or other, of heat or cold, light or shade, love or hatred, pain or pleasure. I never can catch *myself* at any time without a perception and never can observe anything but the perception (Hume 349a-b).

Since Hume denies that there is a self, we can call Hume's theory of self the no-self theory of self.
Whenever a philosophers denies the existence of some thing that people commonly believe exists, s/he has to provide some account of our ordinary beliefs.
If I were to tell you that there is no Santa Claus, I would have to explain to you how the properties you think belong to Santa Claus really belong to other people: your parents bring you presents, a neighbor puts on the Santa suit for the party, the department store hires people to dress as Santa at the mall.
When Berkeley argues that there is no external world, he has to account for our ordinary beliefs in material objects.
Berkeley did that by showing that our ideas of objects could, strictly speaking, be interpreted as about our own sensations; we mis-perceive the world as material.
When the B-theorist argues that past, present, and future are not real, s/he shows how to translate sentences about the past, present, and future into sentences about earlier-than and later-than.

So, when Hume argues that there is no self, he has to provide some account of our ordinary beliefs in the self.
We can evaluate his no-self theory of the self both on the plausibility of his argument for the theory and on the plausibility of his account of our ordinary experiences.

III. The Bundle Theory

Hume's no-self theory relies on his premise that the self must be precisely identical over time.
That claim might seem too strong.
As we age and acquire more experiences, we have different properties.
Even having lunch or shifting our bodies slightly to the left changes our relations to the world without changing our ordinary conceptions of our selves.
The biological theory of the self accommodates these changes by relying on the functional organization of the body as a criterion for identity over time.
Another way of looking at the biological theory would be to see the self as a collection of loosely-related individual instances of selves, each just a moment of time wide.
The self over time would be a bundle of related biological entities.

Hume's account of our ordinary conception of self is similar to this functional view.
Hume argues that we never see a self.
But, we do have experiences.
So, whatever we call ourselves must be related to our series of experiences.
Our experiences are joined by a variety of psychological connections among our ideas: resemblance, contiguity, and cause and effect.
These psychological connections govern all of our thoughts.

They do not connect our ideas in some underlying substance, some haecceity.
These psychological capacities merely conjoin our experiences over time.
Even memory, the essential characteristic of the Lockean self, just demonstrates such mere conjunctions.
The ordinary notion of self which we are pursuing outruns our memories: there are experiences which I call mine that I do not remember.
Locke sees the self as the seat of connections among our memories, among our conscious states through time.
For Hume, there is only a series of loosely-related experiences.

Instead of being a paradigm of unity, Hume thus argues that the self is an exemplar of diversity.
Just as Berkeley argued that an apple is a bundle of independent sense experiences, its taste independent from its roundness and its crunch, we are just a collection of various, independent experiences.
As far as we know, even the world itself is just a loose collection of events, unconnected by causal laws.
Everything is particular, and all the particulars are independent.
Even the self is dissolved.

> When we attribute identity, in an improper sense, to variable or interrupted objects, our mistake is not confined to the expression, but is commonly attended with a fiction, either of something invariable and uninterrupted, or of something mysterious and inexplicable, or at least with a propensity to such fictions. What will suffice to prove this hypothesis to the satisfaction of every fair enquirer, is to show from daily experience and observation, that the objects, which are variable or interrupted, and yet are supposed to continue the same, are such only as consist of a succession of parts, connected together by resemblance, contiguity, or causation... (Hume 350b).

For Hume, then, the self as we ordinarily understand it is just a loose bundle of experiences.
Instead of calling Hume's view the no-self theory, we can call it the bundle theory of self.
That term is a bit misleading, since it might be interpreted as claiming that there is a self which ties the bundle together.
Like the ship of Theseus, we can have a practical interest in maintaining a notion of the self over time.
But, the bundle theorist argues that the claim that there is a self underlying the experiences, some haecceity, is, strictly speaking, false.
There is no single thing which we call the ship of Theseus.
There is no I, beyond the experiences.

IV. Hume, Parfit, and the Buddha

Parfit's view of the self, which relies on some results in contemporary psychology and neuroscience, is another bundle, or no-self, theory.

> In a sense, a Bundle Theorist denies the existence of persons...There are persons or subjects in [a] language-dependent way. If, however, persons are believed to be more than this - to be separately existing things, distinct from our brains and bodies, and the various kinds of mental states and events - the Bundle theorist denies that there are such things (Parfit 352b).

Parfit's claim that there is no self is a contemporary version of Hume's theory.
It is also similar to the Buddhist view of the self, on which there is no self.
On the Buddha/Hume/Parfit view, Descartes's cogito does not yield the existence of a thinker.
There is just thought.

Parfit contrasts the bundle theory with the ego theory, which he attributes to Descartes.
Locke's theory and all of the other reductive theories we have seen are also ego theories.
Even Reid's non-reductive theory of the self is an ego theory.
All of them postulate a subject beneath the experiences.
Parfit argues that it is common to believe in some version of the ego theory.
But such beliefs are false.

> Most of us...have certain beliefs about what is involved in our continued existence over time. And these beliefs would only be justified if something like the Ego Theory was true. Most of us therefore have false beliefs about what persons are, and about ourselves (Parfit 353a).

On the ego theory, there are facts about the self over time.
On the bundle theory, there is no underlying, constant self to support those facts.

Parfit motivates the bundle theory in part by using our intuitions in response to a variety of thought experiments, a method similar to the one used effectively by Locke and Kripke.
He considers teleportation.
Is the person who emerges after a teleportation you, or just a molecule-by-molecule replica of you?
Do you die when you teleport, or survive?
On the ego theory, there should be a determinate answer to this question.
But it seems impossible to discover one.

Parfit considers more-complicated replications.
In these thought experiments, we copy different percentages of our cells.
We can grant that if we replace all of our cells, we have a copy of ourselves.
Still, we can replace smaller percentages: 80%, 60%, 40%, etc.
The resulting person will be less like me the lower the percentage of cells which are replicated.
If none of my cells are replicated, the new person is definitely not me.
There are borderline cases for which the ego theorist runs into difficulties trying to determine which replicas would be me.
The bundle theorist need not make such determinations.

> In the case in the middle of my range, it is an empty question whether the resulting person would be you, or would merely be someone else who is exactly like you (Parfit 354a).

Parfit's claim is that the view that there is a self is the source of the worries about teleportation and replication.
If we rule out the questions as meaningless on the basis of the claim that there is no self, then we can avoid such complications.

> How could it be a real question what would happen to you, unless you are a separately existing Ego, distinct from a brain and body, and the various kinds of mental state and event? If there are no such Egos, there is nothing else to ask a real question about (Parfit 354b).

By framing the argument in terms of thought experiments, one might wonder whether Parfit's defense of the bundle theory relies too heavily on our intuitions.
Perhaps intuitions are not reliable.

Parfit denies that his argument is merely philosophical.

> The truth of the Bundle Theory seems to me, in the widest sense, as much a scientific as a philosophical conclusion (Parfit 355a).

The thought experiments are evocative.
But, the really compelling cases are actual.

V. Split Brains

People who have the left and right hemispheres of their brains disconnected are called split-brain patients.
To disconnect the two sides of the brain, one severs the corpus callosum which connects them.
This is sometimes done voluntarily in persons with severe epilepsy.
Other times, it can be the result of an accident.
In split brain patients, information can be processed subconsciously in surprising ways.
Here is a video of a split-brain patient processing information subconsciously.

One of the remarkable facts about split-brain cases is that persons with split brains can process information in independent ways that persons with ordinary brains can not.
Parfit describes an imaginary example based on the phenomenon shown in the video.
In Parfit's imaginary (but realistic) case, the split-brain patient reports answers to questions by using his hands.
His left hand is controlled by the right hemisphere of his brain and his right hand is controlled by the left hemisphere of his brain.
Similarly, the right brain receives images from the left side of his visual field and the left brain receives images from the right side of the visual field.
Each half of the brain is doing its own processing.
When shown a divided visual field, left-side blue and right-side red, say, and asked how many colors he sees, the patient will respond, with both hands, that he sees only one color.
The right hand will say that the color is blue.
The left hand will say that the color is red.

The ego theorist says that there is a unary thing called the self.
Reid called that thing a monad and argued that we have perfect identity.
In contrast to the ego theorist, Parfit believes that persons with split brains are not monadic.

> In split-brain cases, there are two streams of consciousness (Parfit 352a).

If there are two streams of consciousness in split-brain patients, then such persons can not be seen, on the Lockean view, to be a single person.
There are several possible conclusions that we can make from this datum.
We could conclude that any theory which identifies personhood with conscious experience is wrong; there is one person with a divided consciousness.
Or, we could conclude that such theories of personhood are irrelevant because there is no single stream of consciousness in any one person; there is no person at all.
Or, we could deny that there are two streams of consciousness in the split-brain patient.

VI. The Subconsciousness Reply

Parfit considers the latter response at the beginning of his article.

> It has been claimed by some that there are not *two* streams of consciousness, on the ground that the sub-dominant hemisphere is part of the brain whose functioning involves no consciousness (Parfit 251b-252a).

Notice that Parfit asks the split-brain patient to respond with his hands, not by speaking.
In the video of the split-brain patient, the subject responds confusedly, equivocally, but not in two different ways.
According to the subconsciousness objection, split-brain patients perform subconscious processing which differs from their conscious processing.
People with ordinary brains perform subconscious processing all the time.
But, that doesn't make us multiple persons.
It just shows that our minds are more complicated than Locke and Reid and Descartes thought they were.

Parfit replies that the subconsciousness reply ignores the fact that the sub-dominant hemisphere can actually produce consciousness, if the dominant hemisphere is damaged.
The tasks it performs would be conscious tasks in such cases.
The same areas of the brain light up under fMRIs, which are standard procedures for looking at brain activity.
It would beg the question to deny that processes in sub-dominant hemispheres are conscious in cases in which the dominant hemisphere is functioning normally.

One could maintain that the processing performed by the sub-dominant hemisphere of the brain is not conscious in cases of normally-functioning dominant hemispheres while conscious in cases of malfunctioning dominant hemispheres without contradiction.
Such a response gets us right to the questions of the nature of consciousness and whether consciousness is tied to awareness.
We will return to these questions after break.

VII. Maintaining the Ego Theory

Another way to interpret the relevance of split-brain cases to our discussion of personhood is to accept that there are two streams of consciousness in such persons and to maintain the identification of personhood with consciousness.
This interpretation would thus claim that split-brain cases involve two persons in one body.
This multiple-persons response would violate some conditions on our ordinary notions of personhood.
Despite the Jekyll-and-Hyde-type thought experiments (e.g. Locke's day-night man), we don't really think that one body can house two persons, in real life.
Still, split-brain patients are exceptional.
Persons suffering from dissociative identity disorders, the so-called split-personality cases that are often portrayed in fiction and sometimes found in real life, are also exceptional.
Perhaps we will have to make exceptions to our ordinary beliefs to accommodate such cases.
Neuroscientific research into the brain might give us good reasons to give up some of our ordinary language.
Still, it would be better to maintain as much of our ordinary view as we can.

Parfit denies that there are two persons in split-brain cases, though he does not argue for that denial.

> These cases do not, I have claimed, involve two people sharing a single body (Parfit 355b).

Parfit's reasons might be partially based on the claim that we should maintain our ordinary views as far as possible.
But, since he is defending a bundle theory of self, he is already giving up an ordinary view of personhood.
His broader reason for rejecting the interpretation of split-brain patients as housing two persons in one body is his independent commitment to a more-comprehensive view of personhood: there are no persons.

> Another view is that, in these cases, there are two persons involved...I believe that we should reject this view...[because], in a sense, the number of persons involved is none (Parfit 352a).

According to the bundle theory, there are no persons in the standard, Descartes/Locke/Reid sense, in the sense of the ego theory, as a unary subject of experience.
If we abandon the claim that we have to determine a single person to be the subject of our experience, then we need not worry about split-brain patients undermining our theory of personhood.
Split-brain cases merely support our rejection of the ego theory.

> If there can be subjects of experiences that are not persons, and if in the life of a split-brain patient there are at any time two different subjects of experiences - two different Egos - why should we believe that there really are such things? This does not amount to a refutation. But it seems to me a strong argument against the Ego Theory (Parfit 355b).

Theories are generally considered better when they account for more phenomena.
Newton's theory of gravitation was an improvement on Aristotelian theories of motion because it accounted for both heavenly motions (e.g. planets) and terrestrial motions.
Parfit is arguing that the bundle theory of self accounts for both ordinary brains and split-brains, whereas the ego theory can only account for ordinary brains.
Therefore, we should favor the bundle theory.

VIII. Accounting for the Unity of Consciousness

Above, I mentioned that anyone who denies a commonly-held belief has an intellectual responsibility to account for that belief.
Hume accounted for the denial of persons by appealing to a bundle theory: the self is a collection of independent experiences.
Parfit adds a further requirement on an account of personhood.
On Parfit's view, split-brain patients have multiple streams of consciousness.
We can translate that conclusion to ordinary persons, as well, since the processing in each hemisphere of the split-brain patient is similar to the processing in ordinary brains.
If the split-brain patient has multiple streams of consciousness, we might be seen as having multiple streams as well.
We then have to explain why we seem to have a single stream of awareness.

> There is another explanation of the unity of consciousness, both in ordinary cases and in split-brain cases. It is simply a fact that ordinary people are, at any time, aware of having several different experiences (Parfit 355b).

Parfit's claim is that we falsely believe that we have a single stream of consciousness.
We compile various experiences into what we consider to be a single stream.
Properly speaking, these experiences are independent, like Berkeley's ideas of the apple.

> Just as there can be a single memory of just having had several experiences, such as hearing a bell strike three times, there can be a single state of awareness both of hearing the fourth striking of this bell, and of seeing, at the same time, ravens flying past the bell-tower (Parfit 355b).

It is true that we can experience a variety of different phenomena at the same time.
It is an open question whether there is a single stream of consciousness in which all of these different phenomena are united.
Kant's work, in the late eighteenth century, takes this combination, which he calls the unity of apperception, as the core of his master work, the *Critique of Pure Reason*.
Kant was responding, in part, to the looseness of self found in Hume's work.
But Kant's work itself is widely considered to be ultimately unsuccessful, and questions remain about the nature of the self.

IX. End Personal Identity

We are approaching the end of our work on personal identity.
After break, we will look at a few other cases, and see what to say about them.
Parfit's conclusions are based in real phenomena about split-brain patients.
He also considers, at the end of his article, a case of half-brain transplants, from David Wiggins.
Such cases recall Locke's thought experiments.
Over break, in addition to writing your third papers, you should read about another interesting case, in Daniel Dennett's "Where Am I?".
Then, we will start to look more carefully at the phenomenon of consciousness and the nature of mind and mental states.