Chapter 4c Self
Christopher Shields believes that the solution to the problem of personal identity depends on the proper analysis of the concept of

A. substance  ✓
B. life  ❌
C. consciousness  ❌
D. soul  ❌
Locke’s critique of the immaterial soul

- Locke has established that ships, oaks, and animals can remain self-identical despite having all their matter replaced.
- This gives him an interesting basis for criticizing the view that human self-identity is due to a single, simple immaterial soul.
- His point is extremely interesting (p. 128) : Even if we accept the idea of an immaterial soul, how do we know that a single, simple self-identical substance persists through time?
Locke’s question elaborated

- Remember, Locke has just shown us that a being can be identified with a physical process, where the actual matter is constantly in flux.

- So, why couldn’t a soul just be an analogous immaterial process, where the “soul stuff” is constantly in flux?
Kant’s elastic balls

- The German philosopher Kant imagines a different possibility (p. 129) using an analogy with perfectly elastic balls.
- We know, he says, that when one rolling ball hits another stationery one, it imparts its momentum to it.
- How do we know that the self isn’t like this? Why couldn’t it be that during your lifetime a multiplicity of immaterial substances is simply transferring the contents of your consciousness from one to the other in the way that elastic balls transfer their momentum?
Immortal souls?

- Recall Schopenhauer’s powerful point that our belief in freedom is largely the result of our ignorance of the causes of our actions.

- For those who believe in an immortal soul, Kant and Locke are pointing out that you would never have any way of knowing whether you have the same soul over time.

- In other words, your consciousness could simply be passed from one soul to another and you wouldn’t feel a thing.
Both Kant and Locke believed that the self is due to a single immaterial substance persisting through time.

A. True  
B. False
Human beings vs. persons.

- Interestingly, Locke held that the criteria of identity he developed for oaks and human beings (partaking in the same life) is not sufficient for persons.

- For Locke, 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 different places."

- In other words, a human being is a person to the extent that it is self-conscious, able to reflect on its memories and experiences.
The prince and the cobbler

- Although Blackburn does not discuss it here, Locke arrives at his criterion of identity for persons through a thought experiment.
- He asks us to imagine a prince and a cobbler who exchange souls in their sleep. Hence, the body of the cobbler wakes up in a hovel with all the memories of a prince, and the body of a prince wakes up in a castle with all the memories of a cobbler.
- Locke thinks it is obvious that the cobbler as human being is still in the hovel, but the cobbler as person is now living in a castle and in a prince’s body.
- Hence, Locke argues that the criterion of identity for persons is not a function of it’s physical nature, but of its consciousness. To be the same person through time is to possessed of the same memories.
According to Locke’s view a robot would be a person if

- **X** A. it were made by human beings.
- **X** B. it could speak and act just like other human beings.
- ✔ C. it had the capacity for reflecting on it’s own experiences and memories.
- **X** D. it were made from entirely human parts.
Locke’s motivation

- As a Christian, Locke believes in a soul that survives bodily death; but as a scientist and a philosopher he knew that there is no way to know the soul of another man. Hence, while he puts the thought experiment in terms of changing souls, he does not see personal identity in terms of the soul at all.

- Part of what Locke was concerned about here is the nature of just punishment. He believed (very controversially for the time) that it isn’t just to punish someone for a crime she can’t remember committing.

- Locke’s thought experiment allows him to put his objection in the strongest possible terms: If the person doesn’t remember committing the, then the person didn’t do it, even if his body did.
Locke’s view was groundbreaking, but also highly problematic since it really cannot preserve the idea that we are the same person over time.

The reason is just that from birth to death we forget a lot of stuff. There are all sorts of things you could remember when you were 10 years old that you can’t remember at all now.

For example, on March 27th 1998 you probably could have told me what you had for breakfast the previous morning, but now you’ve almost certainly forgotten.

It’s not clear that this result bothered Locke that much. He may have been happy to conclude that while we are pretty much the same person from day to day, we aren’t really the same person from decade to decade.
Locke’s view implies that you are the same human being from birth to death, but not that you are the same person from birth to death.

A. True
B. False
Thomas Reid (p.131-132) made a famous criticism of Locke’s view with a thought experiment about an officer in the military. Essentially he asked us to imagine this man at three different times in his life: as a young boy, as a military officer, and as an old general. Call these times A, B, and C respectively. Reid asks us to imagine the following things to be true of A, B, and C:

- A: The boy is flogged at school.
- B: The officer captures the enemy’s flag in battle. At this time, he can still remember when he was flogged in school.
- C: The old general still remembers capturing the flag, but has no memory of the flogging.

Reid points out that Locke’s view implies that A=B and B=C, but A ≠ C. This is a problem because it violates the transitivity of identity.
Locke’s defense

Reid’s criticism seems crushing, but there is a fairly simple response to it. Locke can point out that in the real world (as opposed to the purely mathematical one) whether a thing is the same over a certain period of time is going to be a matter of convention.

To explain this point Blackburn asks us to imagine a law that gives Theseus a tax break on his ship as long as it remains physically identical to the one he bought. The government decides that the ship is the same at two different points in time A and B, just in case 55% of the physical material is the same.
Personhood as a forensic notion

- This criterion will also fail transitivity. Suppose:
  - A = newly purchased ship.
  - B = 10 years old, 40% of original material replaced.
  - C = 15 years old, 60% of original material replaced.

- So again, A = B, B = C, but A \(\neq\) C.
- But Blackburn’s point is: who cares? This is not really a mathematical or metaphysical identity, it is just identity for practical legal purposes.
- Locke’s theory of identity has the same defense. It is identity for practical/moral purposes, not mathematical or scientific purposes.
Blackburn’s defense of Locke’s concept of personhood implies that

A. there is really no way to determine whether someone is the same person over time.

B. no human could possibly be the same person over time.

C. most humans are many different persons at the same time.

D. whether we regard a human being as the same person over time is more of a practical matter than a scientific one.
Hume’s bundle theory of the self

- As we saw earlier, Hume argued that the self, considered as a simple entity that ‘owns’ perceptions and experiences is unobservable, and that we should be skeptical of its existence.

- It is often said that Hume subscribed to a different view of the self, namely that it is not the owner of our perceptions, but simply the collection or ‘bundle’ of perceptions itself.

- That’s one way of putting Hume’s view. The other way of putting it, perhaps more charitable, is that Hume didn’t believe in a self at all.
Kant’s critique of Hume

- The main problem with Hume’s view is that it is very difficult for us to make sense of perceptions that don’t have an owner.

- As Blackburn points out, imagining a perception without a perceiver is like trying to imagine a dent without a surface, or a smile without a face.

- Kant puts the point this way. Hume is right that the self is unobservable, and right to reject the self as a simple, immaterial entity that collects experiences in some way that is completely beyond empirical examination.
Where Hume goes wrong

- But Kant thinks Hume is wrong to conclude that the term “I” must therefore refer only to a particular bundle of perceptions.

- In some ways this shows that Hume himself can’t quite get beyond thinking of the self as a kind of thing.

- Kant claims that the word “I” does not refer to a particular thing at all. Rather, it is the expression of a particular point of view.
One way to see what Kant is saying is to adopt what we now sometimes call the “design stance.” What would it be like to design a robot that had some minimal form of self-consciousness?

Focusing on visual experience, Blackburn imagines a robot fitted with a camera. The robot can record visual images on a screen, and in this sense a “bundle of perceptions” is present.

But these perceptions are not of any real use to the robot at this point, because it has no way of knowing what it is looking at.
The self as an organizing principle 2

- For example, if a round shape is moving across the screen, the robot does not know whether it is looking at something small moving slowly nearby, or something larger moving more quickly far away, or even whether the shape itself is moving, rather than the robot.

- The point here is that the bundle of perceptions is not at all useful to the robot unless it has a set of rules for interpreting their significance with respect to it.
In other words, ideas like near, distant, left, right, above, below, before, behind, big, little, fast, slow, etc. all express relations to a particular observer. These relations express the observer’s point of view, and they need to be determined in order for the observer to know how to respond appropriately to their perceptions.

So Kant’s idea is that anytime we refer to an “I” we are not referring to some kind of immaterial substance, or something we could discover in experience, but rather to the set of principles by which we interpret these experiences from a particular perspective.
Kant would agree with Locke that we are not necessarily the same person over time because we simply do not retain all of our memories over time.

A. True

B. False
Other points of view

- If a simple robot like the one described is outfitted with a set of organizing principles it will have a point of view, though it won’t actually know it has a point of view; i.e., it will have a self, but it won’t actually be self-conscious.

- Self-consciousness is partly the recognition that our own point of view can be different from others; because of your particular circumstances you perceive an environment we both occupy differently that I do.

- Blackburn suggests that it is this ability to look at things from another person’s point of view that gives rise to certain delusions, such as that we might survive bodily death.
Delusions

- For example, because you can imagine different points of view, you have the capacity to imagine what it would be like to be at your own funeral.
- You can imagine seeing your body in a coffin, and it is very natural to describe this activity as *your self* seeing your body in a coffin.
- Similarly you can imagine living in a different place or time, which you naturally describe as imagining *your self* living at a different place or time.
- Blackburn’s point here is just that imagining *your self* doing this or that is an expression that shouldn’t be taken too literally. All you are doing is recognizing the existence of points of view besides your own.
Duplicating the soul

- Today philosophers sometimes like to imagine more sophisticated versions of Locke’s thought experiment about the Prince and the Cobbler.

- For example, suppose that we were able to simply upload all of your memories, experiences, skills and aptitudes into the body of another person who looked just like you. When the upload is complete, there will be two different individuals with the same consciousness.

- If we could do this, Locke would have to say (at least at the moment of creation) that two completely different human beings are actually the same person.

- Taking Locke’s view seriously in legal and moral terms is going to make things pretty awkward, as it implies that you both own the same car, both have the same marital obligations, etc.
These thought experiments can also be very confusing.

For example, suppose you woke up in the hospital this morning feeling fantastic, but not knowing how you got there. The doctor tells you that you were in a horrible accident the night before. They weren’t able to save your body, but they were able to copy all of the contents of your brain and upload it into another brain inside a body just like yours, though with none of it’s previous defects. Seems like on the whole you’d find this situation to be pretty agreeable.

On the other hand, suppose exactly the same transfer as above occurs against your will, with the exception that you were not in a horrible accident at all. You wake up to be told that all of the contents of your brain have been uploaded to a new improved body, and that in a few hours they will go ahead and put you to sleep for good. Seems like on the whole you’d find this situation pretty disagreeable.
Past vs. Future

- As Blackburn points out, our intuitions about the significance of these outcomes are very different depending on whether we are looking to the future, or looking back at the past.

- In the first example, looking to the past, you are the surviving improved self, and you are the same person only with an improved body. It’s great.

- In the second example, looking to the future, you are going to be killed so that some other being they say is really the same person as you can live. It totally sucks. But after you are gone, everyone else is just fine with it.

- Of course, we don’t have the ability to perform these operations at this point, and maybe we never will. But these examples do seem to show that our simple notion of a determinate self that either persists fully intact through time, or else completely perishes can not be quite accurate.