Criteria of Identity

We saw that, according to Locke, you can’t just point at two people and ask whether they’re the same one, as the question is ambiguous. You could be asking whether they’re the same *man* (human), or whether they’re the same *person*.

Locke is not the only philosopher to take this view. It is actually quite common to think that one cannot speak about identity *simpliciter* between objects at different times, that in fact talk of such identity is meaningless without some *criterion of identity*.

1. Examples where Criteria are Needed

Here are some examples that help to drive this idea.

*(i) The Ship of Theseus*

Plutarch (*Vita Thesei*, 22-23):

“The ship wherein Theseus and the youth of Athens returned had thirty oars, and was preserved by the Athenians down even to the time of Demetrius Phalereus, for they took away the old planks as they decayed, putting in new and stronger timber in their place, insomuch that this ship became a standing example among the philosophers, for the logical question of things that grow; one side holding that the ship remained the same, and the other contending that it was not the same."

Is this the same ship as the earlier one? There seems to be no determinate answer to this. It’s up to us. We might pretend that it is the same ship, but the facts hardly require this. At most, there is a certain *continuity* in the ship’s history, in that there were never more than a few planks replaced at a time.

(The ship of Theseus example can be made more complicated. Suppose, for example, that when each rotten plank is replaced, an antiquarian takes it, and over time he reassembles these rotten planks, in their original order, into a complete (but unseaworthy) vessel. When he has finished, the antiquarian claims to own the “true” ship of Theseus. But his claim is contested by the owners of the other ship. Who is right? Or are they both right?)

What might a suitable criterion of identity for ships be? If A and B are ships, and B is formed from A by a small modification, then B is the same ship as A? If so, then you can make as many such modifications as you want without losing the original ship. Or: A and B are the same ship if their material constituents overlap by at least 50%?
Note that the first criterion is vague, as it uses the term “small”, which has borderline cases. There will then be borderline cases of identity between ships.

(ii) The Bertrand Russell Society

Suppose five people decide to form the Bertrand Russell Society, for the purpose of studying and promoting the works of Bertrand Russell. They have reading groups, organise public lectures, etc. The Society runs for a few years before folding. Years later, another group of Russell enthusiasts decides to start the Society up again, using the same constitution, doing similar activities, and perhaps even using the same premises. They claim that the Bertrand Russell Society is back, and even apply to the old Society’s bank to have its accounts re-activated. Is it the same society, however, or just a similar one? Is there a fact of the matter here? What would a suitable criterion of identity be?

Most likely, we would again choose some kind of continuity. The Society needs members, and new members have to be approved by a 2/3 majority in a free vote of existing members. In order for a group of people to be the BRS, then, there must exist a continuous chain of votes that connects each one to the five founding members. In that case, the new society is not the same as the old one. If it’s a totally new group, then we’re inclined to say that it’s a different group with the same name, aims, etc. The point, however, is that talk of identity here seems to be meaningless without a specified criterion.

(iii) The Hole

A hole is dug in the ground, remains for a few days, and is then filled in. A year later another hole is dug in the same place. Is it the same hole as the earlier one?

Again, we would likely opt for continuity. While a construction boss might tell a worker “I need you to move that hole over there”, this isn’t really possible. On the other hand, if I dig a hole and a week later someone falls into it, breaking a leg, then I am liable. I might claim that the later hole was not the hole I dug, but the later hole was clearly caused by mine, and this causal continuity seems to be the important thing.

Holes in solid ground are fairly stable entities. Holes (e.g. vortices) in water, on the other hand, can come and go fairly quickly. Two of them can collide, and become one. A single vortex can split into two. There doesn’t seem to be any meaning to identity here, beyond causal continuity.
2. **Criteria of Identity for Abstract Objects**

One of the most common ways to introduce a new abstract object is to give a “criterion of identity” for it. (Another way is to “construct” it out of previously-defined abstract objects.) For example, here’s one way to introduce the natural numbers, i.e. 0, 1, 2, 3, …

Let us stipulate that each well-defined set $A$ of objects has an associated abstract object called its *cardinal number*, written $\#(A)$. This, obviously, doesn’t tell us anything useful. It allows, for example that the sets \{Paris, Rome, Madrid\} and \{Frankfurt\} have the same number, i.e. $\#\{Paris, Rome, Madrid\} = \#\{Frankfurt\}$. To really say what numbers are, we have to give a criterion of identity, i.e. a rule that defines when two sets have the *same* number of elements. This criterion is as follows:

**Definition**

Let $A$ and $B$ be sets. Then $\#(A) = \#(B)$ if and only if there exists a 1-1 correspondence between the members of $A$ and those of $B$.

You can probably see that this definition is all we need, in saying exactly what the natural numbers are.

Here are some other criteria of identity, which are used to make discourse about certain abstract objects meaningful.

1. Two sets are identical just in case they have exactly the same members.
2. Two lines have the same direction just in case they are parallel.
3. Ratios $a/b$ and $c/d$ are identical just in case $a \times d = b \times c$.
4. Two sentences express the same subjective proposition just in case it is impossible to believe them to different degrees.
5. Two sentences express the same objective state of affairs just in case it is impossible for one to be true and the other false.

3. **Identity is an Equivalence Relation**

An equivalence relation is one that is reflexive, symmetric and transitive. Here are the definitions. Note that ‘$aRb$’ means that $a$ is related to $b$ by the relation $R$.

The relation $R$ is **reflexive** iff, for every object $x$, $xRx$.

$R$ is **symmetric** iff $xRy$ entails $yRx$.

$R$ is **transitive** iff $xRy$ and $yRz$ together entail that $xRz$. 
So, for example, ‘x is a brother of y’ is transitive, but not reflexive (Fred isn’t his own brother) or symmetric (Fred might be a brother of Sally, but she isn’t his brother.) The relation ‘x is friends with y’ is symmetric, but not transitive.

Identity has all three properties, making it an equivalence relation. This means that, whenever a criterion of identity is proposed for some object, it is essential to check that the criterion creates an “identity” relation that is reflexive, symmetric and transitive. Any criterion that fails one of these tests must be rejected.

We saw, in discussing Locke’s memory criterion, that it had problems with transitivity, for example.

4. Different Criteria for Personal Identity

The many different criteria for personal identity fall into two basic groups. The first, and most popular, is the psychological criteria, appealing to memory, personality, and other psychological characteristics. To avoid transitivity problems, there is usually an appeal to causal connections. For example, for A to be psychologically continuous with B, one might require that B be caused to have psychological properties by the fact that A has them.

The second main group of criteria are the somatic (or bodily) criteria. If the physical body (especially the brain) is preserved, then one has the same person.

The difference between these criteria is nicely illustrated in a thought experiment by Bernard Williams (“The Self and the Future”, *The Philosophical Review*, Vol. 79, No. 2, Apr., 1970, pp. 161-180). He describes a case where two people, A and B, each have their brains scanned, their memories being copied onto a storage device. Then their memories are erased. Finally, the stored memories are returned to the brains, but the wrong ones! I.e. A’s memories are put into B’s brain, and B’s memories go to A’s brain. Thus when the A-body-person wakes up he will believe he is B, say that his name is ‘B’, and so on. The B-body-person will believe he is A, will try to go “home” to A’s house, recognise A’s wife and children, etc. Each one, when he sees himself in the mirror, will say that he looks different from before, and perhaps even say “I have a new body now”.

Thus the process Williams describes might be characterised as an exchange of bodies between A and B. This is exactly how the memory criterion sees it.

Williams argues that an alternative interpretation of the process is just as plausible, however. Suppose, for example, you are told that you will be tortured tomorrow. Naturally, this makes you very fearful. Then you’re told that, prior to the torture, your memory will be erased. Does this make you sigh with relief? Not at all – if anything, the

---

1 We need to use some such name for this person, since calling him ‘A’ or ‘B’ will beg the question.
scenario is now even more terrifying. Then you’re told that not only will your memory be erased, but that the memories of another person will be implanted in your brain. Thus, while you’re being tortured, you’ll think that you are that other person. Does this make it all ok? Still no, Williams says. The information that you’ll be (effectively) insane at the time, believing yourself to be another person, does not negate the fact that you’ll be tortured. Thus Williams notes that the somatic criterion of personal identity has a lot of plausibility.

5. The Simple (No Criterion) View of Identity

We should not make the mistake of thinking that one has to choose some criterion or other of personal identity, however. One could instead reject the whole business of giving such criteria. The so-called “Simple View” of personal identity does just this, maintaining that two persons are identical just in case they are the same thing. In other words, identity across time is just ordinary identity, a simple matter of being the same substance. On this view, only abstract objects need criteria of identity, for the simple reason that abstract objects have no substance. (The simple view agrees with the somatic view in Williams’ example.)