General Philosophy: Personal Identity

Diachronic and synchronic personal identity

At the most general level, the question to be addressed in this topic is: *What makes one person the same person as some other person?* We can be a bit more precise, though, by distinguishing *synchronic* from *diachronic* versions of this question:

Synchronic personal identity: At a given time, how do I identify a part of the world as a particular person (perhaps myself)?

Diachronic personal identity: What makes two people at two different times, in fact, the *very same* person?

In this topic, we will assume that we know how to identify a particular person at a particular time—that is, we will assume that we can answer the question of synchronic personal identity. The question to be addressed here is that of diachronic personal identity: what makes me the same person as some other person at some other time? In other words: what are a person's *identity conditions* over time?

Before we get to the substance of this question, there are a few more preliminaries regarding the concept of identity which must be addressed.

Qualitative and numerical identity

There are different ways in which two (putatively distinct) objects might be 'the same'. To capture these differences, philosophers often like to distinguish *qualitative identity* from *numerical identity*:

Qualitative identity: Two objects are *qualitative identical* when they have all properties in common.

Numerical identity: Two objects are *numerically identical* when they are, in fact, the very same object.

Example of qualitative identity: 'This ballpoint pen is qualitatively identity to this other ballpoint pen from the same packet'.[1] Example of numerical identity: 'The fastest man in the world circa 2018 is numerically identical with Usain Bolt.'

Philosophers have argued for a long time about how qualitative and numerical identity are related to one another. It seems fairly plausible that numerical identity implies qualitative identity (but see below for some caveats!)—this is known as the *principle of the indiscernibility of identicals*, and can be formalised in (second order!—i.e., quantification over predicates) logic as follows:

$$\forall x \forall y \forall X \left(x = y \rightarrow (Xx \leftrightarrow Xy) \right)$$

This says: 'If x is numerically identical with y, then x and y have exactly the same properties'. Plausibly, though, this only holds *at a particular time*. For I want, plausibly, to say that I'm numerically the same person as that baby so-and-so many years ago—but I certainly don't share all qualitative properties with that baby!

Does qualitative identity imply numerical identity? This is more controversial, and is known as the *principle of the identity of indiscernibles*.[2] This principle can be formalised as follows:

$$\forall x \forall y \forall X \left((Xx \leftrightarrow Xy) \rightarrow x = y \right)$$

This says: 'If x and y have exactly the same properties, then x is numerically identical with y'. The reason that this is controversial is that there seem to be counterexamples: consider, for example, Max Black's *two sphere world*: a possible world with two identical spheres, and nothing else. These two spheres shall all properties (both intrinsic and relational), but nevertheless are distinct. If we think this is a genuine possible world, then the principle of the identity of indiscernibles is false.

[1] NB: Here we're focussing on *intrinsic* properties—the *relational* properties of the pens (e.g., their spatiotemporal distances to other objects) still differ. Also note that, since, presumably, one such pen will have small blemishes which the other will not, the claim in this example is only approximately true.

[2] Physics & Philosophy students will study this in depth on the *Leibniz-Clarke Correspondence* paper.

We don't need to get too far into these muddy waters: the important thing for us to note is that we're concerned with the question of when two people at different times are *numerically idential*—i.e., are in fact the *very same* person.

The memory criterion

One of the three great Early Modern empiricist thinkers, John Locke, proposed the *memory criterion* for diachronic personal identity.[3]

> *[I]n this alone consists personal identity, i.e. the sameness of a rational being; and as far as this consciousness can be extended backwards to any past action or thought, so far reaches the identity of that person ...* (Locke, §9)

In modern parlance, Locke is proposing the following:

Memory criterion: Person P_1 at time t_1 is numerically identical with person P_2 at some later time t_2 iff P_2 can remember what P_1 did at time t_1.

This has some radical consequences. In particular, for Locke, one's bodily constitution is irrelevant to one's personhood—as Locke admits explicitly, in his parable of the pauper and the prince:

> *If the soul of a prince, carrying with it the consciousness of the prince's past life, were to enter and inform the body of a cobbler who has been deserted by his own soul, everyone sees that he would be the same person as the prince, accountable only for the prince's actions ...* (Locke, §15)

Is the memory criterion a plausible account of diachronic personal identity? A different parable—the parable of the general—is sometimes advanced by philosophers wishing to argue against Locke on this front.[4] Here goes:

[3] Remember: the other two great Early Modern empiricists are Hume and Berkeley.
[4] This example is due to Reid.

> A boy grows up, joins the military, becomes a lieutenant, and later a general. The general can remember what the the lieutenant did, and the lieutenant can remember what the boy did. But the general cannot remember what the body did.

Applying the Memory Criterion to this example: the general is the same person as the lieutenant, and the lieutenant is the same person as the boy. Assuming (very plausibly!) that numerical identity is transitive, the general is the same person as the boy. But the general is also, on this account, *not* the same person as the boy, for the former cannot remember what the latter did! Contradiction.

So something seems to have gone wrong with the Memory Criterion. One response to this would be to appeal to the *transitive closure* of the 'remembers' relation.[5] But there are also more direct problems with the memory criterion. Most notably: there are some past times in my life for which I just *cannot remember* what I did. But surely it is to fast to say that I was not *me* at those times? So: it looks like we should seek out some other, better, account of diachronic personal identity.

Psychological continuity

Sometimes people suggest *psychological continuity* as a more sophisticated criterion for diachronic personal identity than the Memory Criterion. Here's how Alex Kaiserman puts the idea in his 2019 lectures on this topic:[6]

> Let's say that P_1 at t_1 is *psychologically connected* to P_2 at t_2 if and only if:
>
> - P_2's psychological state at t_2 is very similar to P_1's psychological state at t_1, and
> - P_2 is in the psychological state she is in at t_2 in large part because of the psychological state P_1 was in at t_1.

[5] 'Transitive closure' means: fill in all the missing arrows to make the original relation transitive.
[6] His slides are available on WebLearn/Canvas.

> And let's say that P_1 at t_1 is *psychologically continuous* with P_2 at t_2 if and only if there is a chain of relations of psychological connectedness leading from P_1 at t_1 to P_2 at t_2 (or vice versa).
>
> The **Psychological Criterion**: P_1 at t_1 is numerically identical with P_2 at t_2 if and only if they are psychologically continuous with one another.

This seems to do better than the memory criterion: it gives the right verdict on the boy/general case, and it also doesn't suggest that I'm not the same person as I once was whenever I forget something. However, the account has some problems of its own:

1. Suppose P_1 is being tortured by P_2, who is herself being tortured. In that case, both of the above bullet points would seem to be satisfied (the first because both people are experiencing the psychological torment of being tortured; the second because it's P_2's psychological state (in particular, her desire to torture) which causes P_1 to be in the psychological state she's in). But we wouldn't want to say that the torturer is the same person as the torturee![7]

2. (A famous example due to Parfit—on whom more below.) It's possible for people who have suffered serious accidents to have one half of their brain irrevocably damaged. However, those people can still survive—and indeed, surprisingly, often their cognitive abilities don't seem to be seriously impacted. It's also the case that (historically in order to treat serious cases of epilepsy) the surgical procedure of *corpus callosotomy* can be enacted, in which the connections between the two sides of the brain are completely severed. Parfit combines these in a thought experiment: suppose you are in a car crash with your identical twin. Your twin, sadly, dies, but his or her body is preserved. The two halves of your brain are severed by a mad doctor, who inserts one half of your brain into your twin's body. Which post-op person is you? Both seem psychologically connected with you, and so to indeed *be* you, by the Psychological Criterion. But surely we can't say they both *are* you! That could lead to puzzling scenarios further down the line—such as your playing tennis with yourself!

[7] I'm grateful to Bryan Cheng for suggesting this objection.

On (2), the idea that there could be more than one thing which is 'you' at a particular time seems wrong—so something would then seem to be amiss with the Psychological Criterion for personal identity, too.

Physical continuity

Williams returns to Locke's pauper/prince example, in an attempt to motivate that diachronic personal identity can't be *merely* a matter of psychological continuity, but must *also* be a function of *physical* continuity.

> *Let us now consider something apparently different. Someone in whose power I am tells me that I am going to be tortured tomorrow. I am frightened, and look forward to tomorrow in great apprehension. He adds that when the time comes, I shall not remember being told that this was going to happen to me, since shortly before the torture something else will be done to me which will make me forget the announcement. This certainly will not cheer me up, since I know perfectly well that I can forget things, and that there is such a thing as indeed being tortured unexpectedly because I had forgotten or been made to forget a prediction of the torture: that will still be a torture which, so long as I do know about the prediction, I look forward to in fear. He then adds that my forgetting the announcement will be only part of a larger process: when the moment of torture comes, I shall not remember any of the things I am now in a position to remember. This does not cheer me up, either, since I can readily conceive of being involved in an accident, for instance, as a result of which I wake up in a completely amnesiac state and also in great pain; that could certainly happen to me, I should not like it to happen to me, nor to know that it was going to happen to me. He now further adds that at the moment of torture I shall not only not remember the things I am now in a position to remember, but will have a different set of impressions of my past, quite different from the memories I now have. I do not think that this would cheer me up, either. For I can at least conceive the possibility, if not the concrete reality, of going completely mad, and thinking perhaps that I am George IV or somebody; and being told that something like that was going to happen to me would have no*

> *tendency to reduce the terror of being told authoritatively that I was going to be tortured, but would merely compound the horror. Nor do I see why I should be put into any better frame of mind by the person in charge adding lastly that the impressions of my past with which I shall be equipped on the eve of torture will exactly fit the past of another person now living, and that indeed I shall acquire these impressions by (for instance) information now in his brain being copied into mine. Fear, surely, would still be the proper reaction: and not because one did not know what was going to happen, but because in one vital respect at least one did know what was going to happen—torture, which one can indeed expect to happen to oneself, and to be preceded by certain mental derangements as well.*
>
> *If this is right, the whole question seems now to be totally mysterious. For what we have just been through is of course merely one side, differently represented, of the transaction which we considered before [i.e., the pauper/prince swap].* (Williams, pp. 167-8)

What's the right conclusion to draw from this? Well, in the pauper/prince case, we thought that personhood swapped—but here we don't—even though it's essentially the same case, just differently presented! If diachronic personal identity were *purely* a function of psychological factors, then we'd think that personhood swapped in both cases. The fact that we don't seems to suggest that personhood is *in fact* a function of both psychological *and physical* factors. As Williams writes:

> *It is also recognized that "mentalistic" considerations (as we may vaguely call them) and considerations of bodily continuity are involved in questions of personal identity (which is not to say that there are mentalistic and bodily criteria of personal identity).* (Williams, p. 179)

One particular group of authors who endorse (at least in part) a 'physical continuity' criterion of personal identity are 'animalists', who say, roughly speaking,[8] that personal identity is a matter of *being the same animal*, where this is ultimately to be cashed out in physical terms.

[8]NB: This is a simplification: see Snowdon's paper for the details.

Super-empirical criteria: the soul

All of the analyses of personal identity which we have considered up to this point have been *empiricist*, in the sense that they've attempted to tie the notion of personal identity to some features of the empirical world (e.g. psychological/physical continuity), to which we can potentially have empirical access. If successful, these analyses would afford us a means of *empirically ascertaining* whether two (putatively distinct) people at different times are, in fact, the very same person.

But ones is not *forced* to adopt a purely empirical criterion of personal identity (whether psychological continuity or physical continuity or some combination of the two), and it may that one's other philosophical commitments open other options. For example: Swinburne is a Christian theistic philosopher, who understands personal identity to be a matter of *having the same soul*.[9]

The problem with this is that, absent those other philosophical reasons to believe in souls, this seems to be metaphysically obscure: explaining something perplexing (personal identity) by way of an *ad hoc* invocation of something *even more* perplexing (the soul). Only if one has other reasons to believe in souls (as, indeed, Swinburne does) is one likely to find this response convincing.

Is personal identity what matters?

In very famous discussion in his masterwork *Reasons and Persons*, Parfit suggests that personal identity "is not what matters". What does this mean? Here's one way of breaking down Parfit's points:

1. Ultimately, 'personal identity'—the notion of 'being the same person'—is a label which we apply to physical systems in the world.

2. We want to use labels which are useful to our practical concerns.

[9] See the Shoemaker/Swinburne exchange for a detailed elaboration of Swinburne's views.

3. What would be such concerns in the split-brain case? Presumably, we would be concerned for the welfare of *both* of our successors.

4. But, as we've already seen, the notion of personal identity (being *you*) can't apply to *both* of your successors—two people at some later time can't *both* be you.

5. So: the notion of personal identity isn't tracking our practical concerns.

6. What *does* track our practical concerns is what Parfit calls the 'R-relation': 'psychological continuity and connectedness, with the right kind of cause'. As we've already seen, both of your post-split descendants *do* stand in this relation to you.

7. So: it is the R-relation which tracks our practical concerns—we should stop being interested in the label of diachronic personal identity, and be interested instead in the R-relation instead.

Some authors, such as Thomas Nagel (*The View from Nowhere*), have resisted this line of argument. Nagel admits that he has the overwhelming intuition that *only one* of his post-split descendants *can be him*, and isn't willing to give up on the project of trying to establish an understanding of the notion of personal identity which allows him to identify *which* one.

References

[1] John Locke, *An Essay Concerning Human Understanding*. **Book II, Ch. XXVII.**

[2] Eric Olson, "Personal Identity", in E. N. Zalta (ed.), *The Stanford Encylopedia of Philosophy*.

[3] Bernard Williams, "The Self and the Future", Philosophical Review 79, pp. 161-180, 1970.

[4] Derek Parfit, "Why Our Identity is Not What Matters", in his *Reasons and Persons*, Oxford: Oxford University Press, 1984. **Ch. 12, pp. 245-280.**

[5] Paul Snowdon, "Persons, Animals, and Ourselves", in C. Gill (ed.), *The Person and the Human Mind*, Oxford: Oxford University Press, 1990. **Ch. 4, pp. 83-107.**

[6] Sydney Shoemaker and Richard Swinburne, *Personal Identity*, Oxford: Blackwell, 1984.