

PERSONAL IDENTITY LECTURE 2

1. Locke takes the prince/tramp thought experiment to show that our identity consists in memory: one is what one is conscious of having been. Distinguish two kinds of memory: factual and autobiographical. I can factually remember that the battle of Hastings took place in 1066; I have an autobiographical memory of having gone to the shops yesterday.
2. There are two well-known objections to this. One is that 'x remembers being y' is not a transitive relation whereas $x = y$ is transitive. The solution is obvious: analyse personal identity as the *transitive closure* or *ancestral* of this relation. If R is a relation then its ancestral R^* holds between x and y if and only if either Rxy or, for some n, there are $z_1, z_2 \dots z_n$ such that $Rxz_1, Rz_1z_2 \dots Rz_{n-1}z_n, Rz_ny$.
3. More interesting is the objection that the criterion is circular. The difficulty is that memory can only be sufficient for personal identity if your apparently autobiographical memories really are memories of what *you* did. We can get round this objection by identifying you with whatever person you inherit most of your *apparently* autobiographical memories (Q-memories) from, or with somebody at the end of a chain of such people.
4. You might also think that a psychological account of personal identity like Locke's must involve other forms of psychological endurance. Parfit's version (RP 216) is more tolerant: you will be whoever is *uniquely psychologically continuous* with you if anyone is. Note the difference, in Parfit's terminology, between continuity and connectedness: psychological connectedness is the holding of particular direct psychological connections, *strong* psychological connectedness is the holding of very many such connections, and psychological *continuity* is the transitive closure of strong psychological connectedness.
5. Williams ('The Self and the Future' in his *Problems of the Self*) considers the following variation on Locke's example. Imagine first that the Prince and the tramp are made aware of what is going to happen to them—in some suitably non-question-begging way. That is, they are told e.g. that the Prince's brain and psychological traits will get transferred to the tramp's body and vice versa (and not e.g. that they will undergo a brain swap, or that they will undergo a body swap).
6. Suppose now that they each get told before the procedure that after the procedure the Prince's body will be tortured and the tramp's body will receive £1M. Presumably Locke would expect the (pre-operation) Prince to welcome this treatment and the pre-operation tramp to dread it. And their respective reactions *after* the operation would confirm this: the person speaking through the tramp's mouth would be glad to have the \$1M and to recollect having looked forward to getting it. And similarly the

PERSONAL IDENTITY LECTURE 2

person speaking through the Prince's mouth would claim that his pre-operation fears (that he would be tortured) were now being realized.

7. Now forget all that and consider the following. It is rational to fear torture in the future. So if somebody whom you believe tells you (a) that your body will be tortured, then you should worry. What if that person then tells you (b) that you will suffer a severe autobiographical memory loss before the torture—of the sort that people sometimes suffer after accidents? That is no reason to be *less* fearful: if anything you now have one *more* thing to worry about. The person now tells you (c) that following the memory-wipe but before the torture you will receive a new set of memories. Again this won't make it any less rational to fear the torture than in cases (a) and (b). The person adds (d) that the memories happen to be those of a now-living individual T; and finally (e) that T will simultaneously get your memories and then \$1M.
8. It is hardly clear that the differences between (a)-(c) and (a)-(d) and between (a)-(d) and (a)-(e), can add up to any difference in its being rational and its being irrational to dread the future torture. The only difference between (a)-(c) and (a)-(d) is that the donor of the memories is still alive; the only difference between (a)-(d) and (a)-(e) is that something similar happens to that donor. But (a)-(e) *is* what happens to the Prince in Locke's story. So it looks as though Locke's story can after all be interpreted as a *mind*-transplant rather than a *body*-transplant. But we now have two wholly incompatible intuitions about the same case. A defender of the psychological view must say that in case (a) by itself you survive in the old body and in case (a)-(d) together you do not, and this can easily seem wrong.
9. Where do we draw the line between levels of psychological degradation that do, and those that do not, lead to death? One might say that there is *no* answer to the question whether you survive in the intermediate cases (compare the stereo). This is very hard to swallow. After all, it seems that if anything has an answer then the question *will you be alive tomorrow* has an answer. Are you then to say that there is simply no answer to the question of when *you* will die? On the other hand what sort of experiment could settle it?
10. Alternatively, we can say that there is always a sharp dividing line between life and death but one that sometimes turns on seemingly 'external' matters, like the difference between (d) and (e), or on trivial ones. Parfit says in favour of the psychological view that even if both alternatives are hard to swallow, we are in any case going to have to accept one of them in his 'combined spectrum' case (*RP* 236-43).