

## Williams on Personal Identity

### 1. The Case

Suppose A and B are going to be put to sleep. Wires will be set up between their brains such that all the information encoded in A's brain will be transferred to B's and vice versa.

When they wake up we'll be confronted with two people:

One person with what-was-A's body and what-was-B's mind. We'll call him *The A-bodied person*  
The second with what-was-B's body and what-was-A's mind. We'll call him *The B-bodied person*

Before being put to sleep it's announced that the one of the two people after the experiment will be given \$100,000 and the other will be tortured.

Each of A and B before the experiment are asked which treatment they want each post-procedure person to get from a purely selfish perspective.

### 2. In Support of Psychological Criteria

Consideration 1: Suppose A chooses that the B-body person gets the pleasant treatment and the A-body person gets the unpleasant person. And B chooses the reverse (that the A-body person get the pleasant treatment).

The experimenter can't act in accordance with both preferences.

Suppose she gives the pleasant experience to the *A-body person* and the unpleasant experience to the *B-body person*, in accord with B's pre-procedure preferences. Then the *A-body person* will be very pleased upon waking up, and claim "I made the right choice! I'm so glad the experimenter decided to go with what I wanted."

Whereas the *B-body person* will be upset and claim: "My choice wasn't honored!"

This is some evidence that the *A-body person* really is B, and the *B-body person* really is A. And that in fact each person made the right choice.

Why is this evidence? Perhaps the thought is:

If X's belief about his past are beliefs about what A did, that is some evidence that X is A.

Consideration 2: Suppose A has anxiety and hopes that as a result of the procedure his anxiety will diminish. After the procedure the A-body person is asked: "Do you feel relief from your anxiety?" The A-body person will respond: "What are you talking about? I've never had any anxiety." If the B-body person is asked the same question he'll say "sadly, this procedure did nothing to relieve my anxiety."

Why is this evidence for the psychological criterion? Perhaps the thought is:

If X's claims about his past mental life are claims about what A's mental life was like that is some evidence that X is A.

More generally, perhaps the thought is: *we should defer to present people who their past selves were.*

### 3. In Support of the Bodily Criterion

Suppose I'm told I'll be tortured. I'll be unhappy about this.

Suppose they say: "while you're being tortured you won't remember being informed of the fact that you'll be tortured"

This won't mitigate my concern

(Note: according to Parfit this should mitigate my concern *a bit*)

Suppose they then add that during the torture I won't remember several things.

This won't mitigate my concern. I'll think torture + amnesia! That's awful!

Suppose they also add that the torture will result in my brain getting messed up in such a way that I have apparent memories that are not my own.

This won't mitigate my concern. I'll think in addition to being tortured I'll have gone mad.

But, Williams says, this is the same as the case described earlier, just presented differently! According to the psychological criteria my comfort should be increasing as the description continues and at the end I should think "this person being tortured will not be me!"

Other intuition-pumps in the vicinity:

- Fears about general anesthesia
- If you were told you'd be in a car accident, and wake up with amnesia and with a different personality, would you want to invest in this person's well being, pain relief, medical care etc? Would you expect your friends and family to care for this person?

### 4. Possible Sources of the Conflicting Intuitions

Possible Source #1: The language, it's described as *I'll* be tortured, *I'll* cease to remember my past, etc.

Response: if we found the idea incoherent we wouldn't/shouldn't go along with it. Williams suggests that, what's underlying the intuition is the thought "my undergoing physical pain in the future is not excluded by any psychological state I may be in at the time" (169) (with certain obvious exceptions).

Possible Source #2: The other person, the person would be *me* in a "body-swap" case, is not even mentioned! So I latch on to the only person in the story who could be me.

### 5. Breaking it Down

(From p.172):

- (i) *A* is subjected to an operation which produces total amnesia;
- (ii) Amnesia is produced in *A* and other interference leads to certain changes in his character

- (iii) Changes in his character are produced, and at the same time illusory ‘memory’ beliefs are induced in him; these are of quite a fictitious kind and do not fit the life of any actual person;
- (iv) The same as (iii) except that both the character traits and the ‘memory’ impressions are designed to be appropriate to another actual person *B*.
- (v) The same as (iv) except that the result is produced by putting the information into *A* from the brain of *B* by a method which leaves *B* the same as he was before;
- (vi) The same happens to *A* as in (v) but *B* is not left the same, since a similar operation is conducted in the reverse direction.

When would you start feeling comforted?

Williams thinks that *A* should clearly be afraid of situation (i), (ii) and (iii).

Doesn’t seem like the fact that the memory impressions are those of an actual human being should make a difference. (iv)

And if we think that why should it matter to me whether they just happened to be the same or whether they were obtained through *B* (v).

So the crucial turning point must happen in (vi), where *A*’s psychology gets transplanted into *B*. But once we’ve gotten to (v), says Williams, why would *A* care what happens to this other person? The move from (v) to (vi), says Williams, seems like one of the least plausible ones to block because the change is a change that happens to somebody else!

Objection 1: You’re begging the question Williams when you say it’s happening to somebody else. On my view it’s happening to *A*!

Response: If you think what happens to the *B*-bodied person in (vi) is happening to *A*, and so in (vi), *A* isn’t being tortured, then you should also think that *A* isn’t being tortured in (v).

After all in both (v) and (vi) the person being tortured – the *A*-bodied person is exactly the same in all respects – character, history, apparent memory. So if you think it’s not *A* being tortured in (vi), you should think it’s not *A* being tortured in (v).

Objection 2: It’s unreasonable to insist that a sharp line be drawn somewhere. It might be indeterminate or ambiguous at which point the *A*-bodied person is no longer *A*.

Response: This may be intellectually comforting but what is *A* who is anticipating all this supposed to make of the fact that in some versions of the case it will be indeterminate whether or not it is *A* being tortured?!

“That it is conceptually undecidable whether it will be me or not, is something which, it seems, I can do nothing with ;because, in particular, it seems to have no comprehensible representation in my expectations and the emotions that go with them” (175).

Analogy: Perhaps it’s like a case in which it’s indeterminate whether an object you’re sentimentally attached to is destroyed?

Response: It’s not. “To regard the prospective pain-sufferer *just* like the transmogrified object of sentiment, and to conceive of my ambivalent distress about his future pain as just like ambivalent distress about some future damage to such an object, is of course to leave him and me clearly distinct from one another, and thus to displace the conceptual shadow from its proper place. I

have to get nearer to him than that. But is there any nearer that I can get to him without expecting his pain? If there is, the analogy has not shown us it. We can certainly not get nearer by expecting as it were *ambivalent* pain; there is no place at all for that. There seems to be an obstinate bafflement to mirroring in my expectations a situation in which is conceptually undecidable whether I occur (177-178).

Williams ends up saying he has no idea what option to choose if he were presented with the original case and that he finds this rather disturbing.

One last consideration: If Parfit is right and we care about psychological connectedness or continuity, then even in cases in which we clearly don't survive, we should still prefer there to be beings who are more psychologically connected to us.

But do we? Imagine a case in which you're killed, but prior to your death some of your memories get transplanted into another person. Does that make you feel *even a bit* better about dying?

Williams, Bernard. "The Self and the Future." *Philosophical Review* 79, no. 2 (1970): 161–80. © Duke University Press. All rights reserved. This content is excluded from our Creative Commons license. For more information, see <https://ocw.mit.edu/help/faq-fair-use/>.

MIT OpenCourseWare  
<https://ocw.mit.edu/>

24.00 Problems of Philosophy  
Fall 2019

For information about citing these materials or our Terms of Use, visit: <https://ocw.mit.edu/terms>.