Williams: A Paradox of Personal Identity

1. **Bernard Williams**: Deceased. Williams had broad interests: he worked in metaphysics, epistemology, and the history of philosophy. His most famous and influential work is in ethics, but we are reading here a contemporary classic on personal identity.

2. **The Problem of Personal Identity**: In virtue of what are you **one and the same** as the person who, 15 years ago, cried to your mother for "extra cuddles"?

   Two ways in which things can be colloquially called "**the same**":

Numerical identity	**Qualitative identity**
"being that very thing, and not another"	"being like and undistinguishable"

   [**EXAMPLE**]: Your ballpoint pen is well-nigh indistinguishable from mine. They are "the same" in every obvious way. But **your** ballpoint pen is not ***this very thing***, no matter how closely resembling it.

   A **numerical identity** claim answers a how-many question; and its denial, a **numerical distinctness** (or **numerical diversity** claim) answers the same question. For instance,

   (1) Marilyn Monroe = Norma Jean Baker

   answers a question like,

   (2) If you put Marilyn Monroe and Norma Jean Baker in a room, how many individuals would there be in the room?

   Or again:

   (3) Joe Biden ≠ the Pope

   answers a question like,

   (4) If you stuck Joe Biden and the Pope into a bag, how many individuals would there be in the bag?

   Evidence: (1) and (3) can be paraphrased by (somewhat awkward) explicitly numerical claims:

   (5) Marilyn Monroe and Norma Jean Baker are one.

   (6) Joe Biden and the Pope are two.

   Compare to the following **qualitative identity** (or **qualitative diversity**) claims:

(7) My pen is the same as A's
(8) My pen is not the same as B's

Unlike (1) and (3), (7) and (8) are not adequately paraphrased by explicitly numerical claims:
(9) My pen and A's are one
(10) My pen and B's are two.

Instead, they say something like "**these things have the same qualities**".

[**BLACKBOARD**]: Draw up equivalences between (1) and (5), (3) and (6), and denial of equivalences between (7) and (9) and (8) and (10).

Questions about identity over time are questions of **numerical identity**.

3. **Personal identity claims have "cash value"**:
The question of identity has three **practical upshots**:

   (a) **Legitimacy of punishment/blame**:
   PUNISHMENT: It is legitimate to punish x for a crime y only if $x =$ the person who committed y

   (b) **Concern for the future**:
   RATIONALITY: If $x = y$, then it is **IR**rational for x **NOT** to be concerned for the happiness of y.

   (c) **Conditions on survival**:
   SURVIVAL: If some event e turns a person x into y, then x survives e if and only if $x = y$.

4. **Two Views of Personal Identity**:

   Psych$_=$ $p_1 = p_2$ if and only if p_1 and p_2 have the same psychology.

   Bio$_=$ $p_1 = p_2$ if and only if p_1 and p_2 "inhabit" the same organism.

   Typical symptoms of having the same psychology: psychological continuity: causal links (of the right sort) between the state of someone's psychology at one time and its state at

another time. Ex: memory connections, carrying out intentions, pursuing goals, satisfying desires, etc.

Typical symptoms of "inhabiting" the same organism: biological continuity: causal links (of the right sort) between someone's biological state at one time and their state at another time. For animals: healing wounds, ingesting and digesting food, moving around, growing, developing, aging, etc.

Williams is a proponent of $\text{Bio}_=$.

These two views are clearly opposed to one another only in cases of alleged body switch or alleged person-switch.

Body-switch cases (*i.e.*, a single person switches bodies) are usually marshalled to support $\text{Psych}_=$.

If body-switch cases are genuinely possible, then they provide an argument for $\text{Psych}_=$ and against $\text{Bio}_=$. Proponents of $\text{Bio}_=$ must argue that, despite appearances, these cases **aren't genuinely possible**.

5. **Williams's Aim**: Williams's aim in "The Self and the Future" is to show that body-switch cases which **seem** at first glance to support $\text{Psych}_=$ are really cases which pose a paradox: they can also support $\text{Bio}_=$.

   Williams's Thesis There are two ways of presenting alleged body-switch cases; on one way of presenting the cases the conclusion that persons have switched bodies seems evidently correct; on the other way of presenting the cases, the conclusion that they have not seems evidently correct.

6. **The original case**:

   (a) **Neutrality**: Williams tries to imagine a way of describing a case in which memories and other psychological features of one person are transfered into another body (and *vice versa*) which is **neutral** on the question of whether psychological or bodily continuity underlies personal identity.

   (b) **Technology**: Williams imagines that the memories and other psychological features of a person can be "downloaded" into a machine, and then "uploaded" into someone's brain.

(c) **The Case, Neutrally Described**: A and B are in a lab equipped with the requisite technology. A's psychology is "downloaded" into machine M_A, and B's psychology is "downloaded" into machine M_B. Then the psychology stored in M_B is "uploaded" into the brain of A's body and *vice versa*.
[**BLACKBOARD**]: **DRAW THE CARTOON**.

(d) **Terminology**:

A-body-person The person who inhabits A's old body after the procedure.

B-body-person The person who inhabits B's old body after the procedure.

(e) **Two Crucial Facts**:

i. The *B-body-person* is psychologically continuous with A;

ii. The *B-body-person* does not have the same body as A.

(f) **Two Preliminary Conclusions**:

i. If Psych$_=$ is true, then the B-body-person $= A$ and the B-body-person $\neq B$.

ii. If Bio$_=$ is true, then the B-body-person $\neq A$ and the B-body-person $= B$.

7. **Arguments for Psych$_=$**:

(a) **Assumption**: Either Bio$_=$ or Psych$_=$ are true. (After all, **what's the alternative**?)

(b) **The Generic Argument**:

i. A and B have switched bodies: the B-body-person $= A$ and the B-body-person $\neq B$.

ii. If A and B have switched bodies, then Bio$_=$ is false. [by Prelim. Conc. (i)]

iii. If Bio$_=$ is false, then Psych$_=$ is true. [By our assumption above]

---

iv. Psych$_=$ is true.

(c) **The whole dispute turns on whether A and B have switched bodies** This argument relies on premise (i). Obviously, a proponent of Bio$_=$ would reject this premise. Why think it's true?

(d) **The argument from memory**: The B-body-person apparently remembers things that A did, but not things that B did. The B-body-person displays the character traits of A and not B. Thus, the B-body-person displays all of the normal symptoms of being A (in disguise, perhaps), and none of the normal symptoms of being B.

(e) **The argument from hope (and fear)**: Suppose the scientists told A about the operations. And they also told him that afterwards they would give one of the people \$100,000 and would torture the other. Should A hope that the A-body-person or the B-body-person gets the \$100,000? A should hope that the B-body-person get the money.

The argument seems to be something like this:

i. It is rational for A to hope that the B-body-person gets \$100,000 rather than torture only if $A =$ the B-body-person.

ii. It is rational for A to hope that the B-body-person gets \$100,000.

---

iii. $A =$ the B-body-person.

Notice that the first premise of this argument bears a significant resemblance to (RATIONALITY).

8. **Williams's Redescription**: Now consider the following description of a situation:

> Someone in whose power I am tells me that I am going to be torured tomorrow. [. . .] [W]hen the moment of torture comes, I shall not remember any of the things I am now in a position to remember. [. . .] He now further adds that at the moment of torture . . . [I will] have a different set of impressions of my past . . . the impressions of my past with which I shall be equipped on the eve of turture 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 on 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. (pp. 167-8)

9. **Williams's "Slide"**: **Williams's Strategy**: start out with a case in which fear is **clearly justified**; add incrementally add to the description of the case, and ask what the reaction is supposed to be.

   In each of (i) through (vi), the treatment is followed by torture (p. 172):

	Case	**Appropriate Reaction**
(i)	Total Amnesia induced in A	FEAR
(ii)	(i) + A's character is changed	FEAR
(iii)	(ii) + A gets illusory memories	FEAR
(iv)	(iii) + A's new psych. fits B	FEAR
(v)	(iv) + B is unaffected	FEAR
(vi)	(iv) + B gets A's psychology	FEAR

10. **NOTICE**: Case (vi) just is the original case, described differently. That is, we have two descriptions of the same case:

    The Neutral Description , stated in terms of the A-body-person and the B-body-person; and

    The A-Centered Description , stated in terms of what is going to happen to A.

11. **Williams's Claim**: It is rational for A to be concerned (indeed, terrified) that **he himself** will undergo the torture ad-

ministered to the A-body-person in case (vi).

12. **An Argument against Psych$_=$**:

    (a) It is rational for A to fear that the A-body-person will be tortured only if A = the A-body-person.

    (b) **Williams's Claim**: It is rational for A to fear that the A-body-person will be tortured.

    (c) A = the A-body-person.

13. **the state of the debate** according to Williams:

    - When presented with **the Neutral Description**, our intuition is that it is reasonable for A to have concern for the future of the B-body-person. This motivates the adoption of Psych$_=$.
    - When presented with **the A-Centered Description**, our intuition is that it is reasonable to A to have concern for the future of the A-body-person. This motivates the rejection of Psych$_=$.

    THE UPSHOT: Those intuitions which have been taken to motivate Psych$_=$ should not be taken as authoritative. **The argument for Psych$_=$ from body-switch cases is not supported**.

14. **Does Non-Neutrality Matter?**
    OBJECTION: William's non-neutral description skews our judgment irrationally.

    (Compare: government-provided assistance polls much worse when they are called "handouts" or "welfare". This is called a "framing effect.")

    (a) Imagine that **you** are A, and the scientists are describing what will happen to **you**:

> [D]oes his use of the second person have a merely emotional and rhetorical effect on me, making me afraid when further reflection would have shown that I had no reason to be? It is certainly not obviously so. The problem just is that through every step of his predictions I seem to be able to follow him successfully. And if I reflect on whether what he has said gives me grounds for fearing that I shall be tortured, I could consider that behind my fears lies some principle such as this: that my undergoing physical pain in the future is not excluded by any psychological state I may be in at the time, with the platitudinous exception of those psychological states which in themselves exclude experiencing pain. In particular, what impression I have about the past will not have any effect on whether I undergo the pain or not. (p. 169)

(b) **Williams's Principle**: Undergoing physical pain in the future does not require that the victim be in any particular psychological state (other than pain) in the future.

Cases:

- Certain unpleasant surgical procedures
- Temporarily altering your memories, beliefs, *etc.*, to fit someone else before causing you pain.

That's what makes the fear rational.

(c) **NOTICE**: Williams makes a big deal out of the fact that the description of the case is "first-personal". You imagine that **you yourself are** A, and the scientists are saying what will be done to **you**. But it seems to me as if switching it to the third person, imagining something happening to A and asking what **he** should fear, makes no difference.

15. **Williams's Challenge**: Figure out where in (i) - (vi) to break the "slide". At what point is FEAR no longer the appropriate reaction?

16. **Meeting Williams's Challenge**: It seems to me that a defender of Psych$_=$ could argue:

(a) **The "slide" should stop after (iii)**: A has been killed before the torture begins; but

(b) **There's another reason for fear**: Fear is the right reaction, not because of impending torture, but because of impending death. (Together with the horror of being in the clutches of someone who would do such a thing.)

[**BLACKBOARD**]: draw a skull-and-crossbones at (iii).

17. **PROBLEM**: The defender of Psych$_=$ is **committed** to saying that A survives (in B's body) in case (vi). This means: The defender of Psych$_=$ should treat treat cases (v) and (vi) differently.

    But why should what happens to **somebody else** determine whether you survive?

18. **Two Difficulties for Psych$_=$**:

    (a) It is implausible to think that whether you survived a certain change depends on what happens **in the future**.
    CASE: We can't tell whether the scientist is guilty of murder or not, because the victim will survive if the scientist decides to upload the person's psychology to a single person.

    (b) It is implausible to think that whether you survived a certain change depends on what happens **to someone else**.
    CASE: We can't tell whether the scientist is guilty of murder or not, because we don't yet know what the scientist did to B.

19. **Williams's Objection** to Psych$_=$:
    As we have seen, Williams gives us reason to think that we can't rely on the body-switch intuitions as part of an argument for Psych$_=$. So we have no good reason to adopt Psych$_=$.

    But Williams thinks he **can do better**: we also have reason to reject Psych$_=$.

    (There is a difference. **Compare**: I've got no reason to think you secretly have a copy of Michael Jackson's *Thriller* in your closet. But I've got no reason to think you **don't** either. Likewise, I've got no reason to think you failed to turn in your paper. But I've got plenty of reason to think you did not fail to turn in your paper.)

> By the sorts of methods [the scientist] employed, he could easily have left off earlier or gone on further. He could have stopped at situation (v), leaving B as he was; or he could have gone on and produced two persons each with A-like character and memories, as well as one or two with B-like characteristics. If he had done either of those, we should have been in yet greater difficulty about what to say; he just chose to make it as easy as possible for us to find something to say. [Omit aspersions cast on the soul theory.] [The scientist] has just produced the one situation out of a range of equally possible situations which we should be most disposed to call a change of bodies. As against this, the principle that one's fears can extend to future pain whatever psychological changes precede it seems positively straightforward. (pp. 179-80)

The idea: cases of person-fission show that there is something wrong with Psych$_=$, **even when there's no fission**.

[**BLACKBOARD**]: draw two fission and one switch cartoon side-by-side.

But why does the **possibility** of other operations affect what's correct about the **original case**?

Problem: the question of survival (and of personal identity) is made **too fragile** by Psych$_=$: According to Psych$_=$, A's identity with the B-body person **depends on what happens elsewhere**.

Cases:

- **"It depends on what's happening in Timbuktu."**: Suppose the scientist attempts two uploads: one here and another in Timbuktu. Whether the person in room #1 = A depends on the success or failure of an upload in Timbuktu.
- **"It's too soon to tell"**: Suppose the scientist keeps a copy of the download. Whether the person in room #1 = A depends on what happens in the distant future. After the person in room #1 dies (of natural causes), whether he *was* A depends on whether there is another successful upload.

Locality of Identity The numerical identity or distinctness of a person x and a person y does not depend on what happens to people at other places and times.

Objection to Psych$_=$:

(a) **Locality of Identity**: The numerical identity or distinctness of a person x and a person y does not depend on what happens to people at other places and times.

(b) If Psych$_=$ is true, then the numerical identity or distinctness of A and the B-body-person depends on whether A_2 is created elsewhere.

(c) Psych$_=$ is false.

Another objection to Psych$_=$: It is incoherent in the fission cases:

(a) Psych$_=$ is true. [ass. for *reductio*]

(b) A is psychologically continuous with each of A_1 and A_2. [description of the case]

(c) If Psych$_=$ is true, then $A_1 = A$ and $A = A_2$. [by (b) , application of Psych$_=$]

(d) $A_1 \neq A_2$ [Premise]

(e) $A_1 = A$ and $A = A_2$. [(a) + (c)]

(f) $A_1 = A_2$ [by (e) + transitivity of =]

(g) Psych$_=$ is false. [by (a) + (d) + (f)]