Bernard Williams: The self and the future - An Analysis and Critique

Introduction

An initial Question: Is it possible that you could exchange bodies with someone? You are to understand 'possible' not as 'practical' or as 'possible', but only as something that is coherent or possible in thought. If you think it is possible, describe the circumstances that would have led for this to happen. And if you think it impossible, say why.

This question is raised preparatory to discussing a thought experiment described in Williams' paper. A similar idea, to which Williams is reacting, occurred in the late seventeenth century in Chapter 27 of John Locke's Essay Concerning Human Understanding, in which Locke imagines a Prince's consciousness inhabiting the body of a cobbler.

Technicalities aside (a wiring problem) there seems nothing conceptually incoherent about having one's own brain transplanted into another's body. The alternative is the "dump and restore" technique - where the contents of one's brain (one's memories and presumably one's capabilities) are copied onto a storage device and then further copied onto the physical brain of another, overwriting or erasing whatever was there beforehand.

There seems nothing utterly inconceivable about either of these techniques, but they seem to be conceptually different. Williams recognises that describing such thought experiments as "changing bodies" begs the question of what constitutes the substrate of consciousness and what makes up an individual person.

The "brain transplant" approach has the benefit of some physical continuity, as the brain is a physical thing, and we are reasonably confident that the transplanted brain would be capable of retaining its information processing capacities. It would have a new body to get used to, but we would expect the mental abilities and memories of the transplantee to be as they were beforehand, though the physical abilities would be much changed and mental functioning would be affected by a different hormonal and respiratory environment and by a different sensory apparatus. Whether the hybrid felt himself to be the same person after this procedure would depend on that person's self image and capabilities. A man whose abilities and aspirations were predominately physical (such as a boxer or footballer or even a musician or painter) would presumably not feel himself the same person after the operation, because it is not likely that these capabilities would cross over successfully, though a philosopher might not consider himself so badly off. Saying you have "exchanged bodies" while retaining your personhood would imply that your body was not of much significance to you as a person other than a rather passive vehicle for your mind. Not many people can say this.

The "dump and restore" approach leaves the added complication of possibly cloning the original person (or at least their consciousness) - the original "me" can be left
there undisturbed, after the contents of my brain have been copied, to awake as before. There is no necessity to copy the other person's brain-contents into my brain. I don't feel fully confident that the "restored" me would actually wake up - maybe it would be a zombie - though presumably a non-vitalist should have the courage of his convictions. Life would presumably be even harder for the new me in this "asymmetrical copy" case than in the "brain transplant" situation, as I would not only have the rest of the world confusing me with the previous owner of the body I then occupy, but I would have to contend with someone with a prior claim to being me - though as he is in on the act, this might be more of a help than a hindrance. It is this asymmetrical situation, which Williams turns to right at the end of his essay, that raises the most interesting questions.

Enough of this pre-amble … we'll see how these initial prejudices (admittedly already influenced by Williams) change after some serious consideration of Williams' thought experiment. We now analyse in some detail what Williams has to say in his essay.

**Analysis of Williams' Paper**

Williams asks us to consider his (tendentiously-entitled) "exchanging bodies" experiment. We have before us two persons, A and B, each with their own memories and characters and distinctive physical mannerisms for displaying them. After the exchange, however achieved, body A displays the memories and character of person B by physical manifestations appropriate to person B, and vice versa.

Williams adopts a simplifying assumption to make the experiment more likely to be understood as that of "exchanging bodies" - he assumes some similarity between A and B so that there is a good chance of B's mannerisms in A's body not being made unrecognisable by gross physical or psychological differences between A and B.

For us seriously to be convinced by the B-ishness of A-body's post experimental character, another condition must obtain according to Williams. B's memories must be reflected by A's body in such a way as it seems true that there is a causal chain between B's experiences and A-body's demonstration of apparent first-hand experience of them (presumably to guard against the supposition that A had simply studied B's background and the "memories" are learned rather than records of first hand experience). Williams insists that the causal chain should not run outside of A's body, and therefore the simplest way to ensure this is by transplanting B's brain into A's body. However, Williams thinks a less radical version of copying B's memories and then restoring them to A's brain will suffice, for the reason that of the three grounds for knowing about one's own past - remembering, being reminded and learning again - there is no sufficient reason for describing the restored memories of B in A as "learned again". It is an interesting question whether this will do, because for A-body to be B, he needs not only B's memories, but B's information processing abilities. It is an open question whether the contents of B's brain can be copied onto different hardware (or in this case "wetware") while retaining the B-ishness of B. The restore mechanism would need to reconstruct the entire structure of A's brain to be as B's (which it might have to do in order to accomplish memory transfer, as memories
are (most likely) physically encoded as connections between neurons). However, the firing rates of these neurons would presumably remain A's. Hence, I prefer the brain-exchange as the most likely method of "exchanging bodies". This form of the experiment is not, however, as interesting.

After this exchange, we have the A-body-person and the B-body-person (where the A-body-person is the person occupying the body that was A's prior to the experiment). Those unaware of the experiment will initially presume the A-body-person is A, while the description of the process as "exchanging bodies" presumes that the A-body-person is B. A non-question-begging approach leaves it open as to whether either the A- or B-body-person is A.

Williams now tries to determine which person is which using a thought experiment from the third-person perspective. A & B are informed before the procedure that post-operatively one of A- and B-body-persons will be tortured and the other given $100,000. They are asked what, on selfish grounds, they would prefer to happen to which. Williams notes that, depending on the choices of the victims, the experimenter may or may not be able to satisfy both of them, but that their choices and reactions to being told prospectively what is going to happen will reveal how they understand the procedure to work in accord with personhood. As Williams notes, while if the post-operative state is announced beforehand it makes sense to say prospectively that A or B got what he wanted, it is an open question whether retrospectively either of them can be said to have got what he wanted, as this begs the question that either of the A- or B-body-persons is A or B, or whether, for instance, they have been hybridised.

Williams suggests that there are good grounds for presuming that we could say retrospectively that either A or B got what he wanted. He takes the case where A and B presume this to be an "exchanging bodies" experiment, so that A chooses that the B-body-person receives the good treatment and B the A-body-person. He then further supposes that the experimenter in fact awards the good treatment to A and the bad to B, though he doesn't tell them beforehand. After the experiment, the A-body-person, with B's memories, remembers that he chose the good outcome and is doubly satisfied on the grounds both that he got what he wanted and that what he got was good. The B-body-person, with A's memories, is similarly doubly dissatisfied. At face value, this seems to imply that B chose wisely, got what he wanted, and that the A-body-person really is B, whereas A also chose wisely, but was unlucky, and that the B-body-person really is A - and that "exchanging bodies" is a correct way of describing the experiment.

That the A-body-person really is B and that the B-body-person really is A can be made to seem probable by variations on the experiment - eg. with the same experimental outcome as before, A chooses that the A-body-person receives the good treatment and that consequently the B-body-person is unhappy with the outcome, but acknowledges that this is what he chose and that he chose unwisely (and similarly, mutatis mutandis, for the B-body-person). Finally, Williams considers the hybrid case where A chooses unwisely and B wisely, with the experimenter acting as before. Both get what they chose, but B-body-person is happy, A-body-person unhappy with the outcome. All three cases seem to support the "exchanging bodies" hypothesis.
Williams now starts to expand the perspective of the experiment by considering the post-operative responses of A- and B-body-persons to their bodies (he is careful not to describe them as B & A respectively, though notes that their responses would be consistent with this assumption - he also notes that if they had viewed the experiment as changing bodies, they would have had to be reasonably satisfied with these bodies before agreeing to take part in the experiment at all). He tries to tease apart the consequences of the supposed body-swap by asking questions such as, if the B-body had a wooden leg that B had become habituated to, would the post-operative B-body-person also be habituated to it; i.e. is the habituation a mental or a physical thing?

Williams now proceeds to consider the matter laterally by thinking in detail about A & B's psychological expectations, concerns and responses. In particular, he stops begging the question by presuming that the experiment results in A and B exchanging bodies. It thus at least makes sense for A and B to ask whether, post operatively, they may be able to escape some of the mental hang-ups they currently have. So, he imagines A with bad anxiety and B with fearful memories, but concludes that the post-operative B-body-person would possess this proneness to anxiety while the A-body-person would have the bad memories. This is still consistent with the "exchanging bodies" hypothesis, and so far the argument leads us to conclude that the requirement for bodily continuity as a necessary condition for personal identity is mistaken and that, as Locke thought, we should identify ourselves with our memories rather than our bodies.

Williams now considers what appears to be a different thought experiment, but which is really the previous experiment this time viewed from a first-person perspective. I am told that tomorrow I am to be tortured, but with allegedly increasingly ameliorating circumstances, as below:

- Shortly before the torture I will forget that I have been forewarned
- I will lose all my former memories
- I will be given a new set of memories, i.e. a new past
- That my then memories will correspond to those of another person now living

The point is that, at no point in this Sorites-type argument will I feel any comfort; rather, I will feel even more disquieted - for (despite being told how this is to be brought about) not only would I still have the prospect of torture to endure, but would have mental derangements equivalent to total amnesia and madness imposed on me as well.

As this is just another view of the first experiment, can we confidently say this view, rather than the first, in which I would be happily escaping from the soon-to-be-tortured body, is wrong. Williams thinks not. The experimenter may simply be intimidating me by persistently using the term "you" to me - thereby begging the question in the opposite direction to the "exchanging bodies" perspective - but it seems clear that, at the moment of torture, whatever impressions I have of the past
will not influence my then present pain, and in reviewing the process there seems to be no point at which I've been "beamed up" into another, happier body.

Williams now briefly explains why he's chosen the example of torture as a future dread event, rather than some other thing one might fear. This is because many of our fears are character- or memory-based, which, in this experiment, are likely or know to be about to change. However, aversion to physical pain is minimally character- or belief-dependent. Having started on this aside, Williams also points out that it may still be valid to fear a psychological disturbance in which our then selves would be perfectly happy - as in our being turned into contented vegetables - and the reason we would fear such a turn of events would be selfish rather than the altruistic acknowledgement that we'd be unable to fulfil our obligations. Personally, I do not view this as a paradox of hedonistic utilitarianism (nor, I suspect, does Williams). What makes our pleasures so exquisite is that we have chosen to have them, maybe struggled for them, and so the utility of a life of our own, with all its vicissitudes, exceeds that of one in the orgasmatron.

Returning to the chase, Williams now points out the second difference between this first-person report of the events and the "exchanging bodies" one - there is no mention of the second person, other than as the source of my new memories. From this first-person perspective, this second person is irrelevant except as an object of our envy, but in the third person account this is the new me, the one on whose account I ought to be afraid, if at all. One who subscribes to the "exchanging bodies" interpretation of events will count this as a fatal objection to the first-person account. However, Williams doubts this is so.

To demonstrate why not, Williams rehearses the first-person account again, imagining the torture occurring at the end of each of the six steps in the process (starting with me suffering total amnesia, and ending with the other person undergoing the analogous character-change to that which I undergo). Since these steps are important for subsequent discussion, and already succinctly summarised by Williams, I have filched them verbatim from Williams' paper :-

(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 certain illusory 'memory' beliefs are induced in him: these are of a quite 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;

---

1 This allusion to Woody Allen's *Sleeper* was made before the recent announcement of the invention of such a device, honest!
(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.

At the prospect of none of these six cumulative experiments do I feel anything other than disquiet. This is because from this perspective at no stage do "I" (A) escape into the other person's (B) body.

Viewing everything up to the fifth step (copying B's dispositions & memories into A's brain while leaving B otherwise alone), Williams thinks we have two answers to the "exchanging bodies" objector. Firstly, there is no reason (given the primal pain-aversion common to us all) why I should not be just as afraid of torture, even with B's dispositions, as normal. Secondly, because B still exists undisturbed, A-body-person (me) cannot be (numerically) the same person as B as there are two persons with a claim to B-ness. Locke would say there are two men but only one person. I would agree with this, but only for an instant - the two men diverge into two persons as soon as either has an experience (I wanted to say "has an experience not shared by the other" but feel that even if their experiences were miraculously kept synchronised they would still be two persons because they are two consciousnesses). From the "exchanging bodies" perspective, if we stop after step five, A has died (or is at least in suspended animation) and B has been partially cloned. This situation is critical. B's consciousness cannot have "hopped" to body-A as it must still be in Body-B at this stage. So, what consciousness is in Body-A? It is tempting to think of this consciousness as a scrambled version of A's, but we "exchanging bodies" types should maybe stick to our guns and assert that the consciousness in Body-A is also B's; for the consciousness in Body-A it will feel like B having swapped bodies, and having temporarily a mental twin in Body-B, though because of its necessarily different environment it will rapidly diverge from Body-B's. Williams assertion that at the end of step five A-Body-person and B-Body-person are certainly not the same person is too strong (though, as I noted, they soon would be).

Williams now tries to find out whether we can make out that the transition from (v) to (vi) is important for A, and does not simply refer to something happening to someone else (B). The point is that, at the end of (v), B-Body-person is very definitely B, so if A-Body-person is not A, then no-one is. This seems fair enough (as I've alluded to above), but if A doesn't exist at (v), when in the process does he disappear? Williams argues that A still existed after (i) and (ii), that is, after the total amnesia and personality change, so maybe we should draw the line after (iii) or (iv), memory exchange with a fictitious person or, in (iv) B. Williams question is rhetorical, presuming a "no" answer. I think, though, that he gives insufficient weight to the catastrophic nature of the changes affecting A-body-person. His analogy is with the normal amnesia, personality disorders or delusory memories that might afflict someone going out of his mind. Well, if someone had been afflicted to the total degree suggested in the thought experiment, we would forensically count them as having become a new person, and treat the original person at least as being in abeyance. The question is, when would A suddenly suffer a dislocation of consciousness, so that a new person with no historic conscious continuity with A pop up in A's body. Personally, I think this would be after step (i), the total amnesia, though possibly after (ii), the total personality change. We could apply more granular Sorites-type
arguments against this (memory draining away bit by bit, personality changing gradually) and ask just when does A cease to exist. This is not the same situation, however - A would be adjusting to this unfortunate state of affairs and would evolve into A*. We wouldn't be troubled by such thoughts in the "brain transplant" version of the experiment. If A's brain was taken out of A's body, there would be no chance of A's consciousness remaining there, not even in his little finger²!

Williams now repels a rather silly proposal, along the lines that we can't decide whether or at what point A-body-person ceases to be A, so why worry about it; some things are just like that. Well, Williams adopts the first person perspective of A, and points out that, while this situation might be acceptable for a third party, it is of vital interest to A, who is either going to be tortured or not depending on the outcome. I have to say I can almost feel Williams being argued into accepting Pascal's wager (after all, what is worse than being tortured for ever !), so there must be something dodgy going on here. Williams labours the rather obvious point that I will retain fear at the prospect unless I am sure that I won't be involved in the unpleasant things yet to happen to A-body-person, and will be fearful in proportion to the probabilities involved. Williams seems to be suggesting that in this case, the situation must resolve itself one way or another - the coin has to be heads or tails - A-body-person will either be me or not, but until I know which, I have good reason to be in trepidation.

Williams thinks that this situation of undecidability is unthinkable from a first person perspective in that if I lose my fear on account of the undecidability, I have effectively decided that it will not be me who will be tortured, while if I continue to worry about it, it is because I think it will be me. Williams tries to envisage whether I might adopt ambivalent concern towards the event, as I might towards something to which I was sentimentally attached that underwent some puzzling confusion of identity. Williams, of course, thinks this won't do, for as soon as I adopt this ambivalent stance towards A-body-person, while I may be hazy about who he is, I've already concluded that he's not me. If I still thought he might be me, I wouldn't be so detached. I think that Williams is getting into a muddle here, and multiplying zero by infinity and getting any number he likes (as in Pascal's wager). If the prospect was a remote chance of a slapped wrist rather than unbearable torture one would easily become dispassionately involved.

Williams states that "there seems to be an obstinate bafflement to mirroring in my expectations a situation in which it is conceptually undecidable whether I occur". Isn't this parallel to any future contingency ? What's the difference between this and being worried on being told that all first-year BA students who fail their exams will be mercilessly tortured. While I might concur that this sentence, though just, is unlikely to befall me - I would do well to fear it because, confident though I may be, there is no way of knowing that I will pass.

Later, Williams considers, returning to the six-stage experiment, whether we might not choose to identify A with A-body-person after stage (v) because there is no better candidate, but not after (vi) because B-body-person will then do much better.

² An allusion to another thought experiment by Locke.
Williams thinks this is like disposing of the effects of some intestate relative - we just have to decide as best we can within the confines of the law. Williams rightly doesn't think this will help A at all, for if he's still frightened at the prospect of (v), the thought that there would be a better candidate for A-ness after (vi) will not console him.

When Williams starts to sum up, he thinks the opposite conclusions reached by the third and first person perspectives of A's fate are conceptually undecidable, and that he's not sure which choice he'd make were he to have the bad luck to be A. Not surprisingly, he's disturbed by this (he's taking his thought experiment rather too seriously, one might think). Williams brings up the dichotomies between, respectively, first and third person perspectives and "mentalistc" and "bodily continuity" considerations involved in questions of personal identity. Williams points out that his thought-experiments have revealed a reverse parallelism to that usually considered for these two pairs of concepts - the third person perspective is here associated with "mentalistc" considerations while adopting a first person perspective led us to support bodily continuity. Williams considers this inversion of some enigmatic significance.

Finally, Williams considers whether the presentation of the third-person perspective in its neat symmetrical form unfairly induced us to consider it as "exchanging bodies". This is a very important point, and in my view more space should have been devoted to it at the expense of the angst-ridden ramblings that occupy most of the latter half of the paper. Clearly, according to the rules of the experiment, we could have generated any number of A's and B's in any number of brains / bodies. This is in sharp contrast to the more invasive, but allegedly equivalent, "brain swap" alternative experiment. This leads me to feel there is a little bit of slight of hand in Williams' introduction of his non-invasive equivalent, which is not equivalent in the least - though depending on what Williams is seeking to demonstrate, this lack of equivalence may not matter. The possibility of "me" being "restored" into the brains of numerous bodies (while being left alone in my own body - let the reader please forgive the tendentious language here !) makes it unlikely that my consciousness would flip over into B-body-person at any stage in the experiment - otherwise, in variants of the experiment, I could be saddled with an open-ended number of (presumably incommunicable) selves to cope with. We would have to suppose that these new consciousnesses, for all their similarities to me, are not me, and that in the experiment both A and B die, with two new consciousness emerging in their places.

So, Williams comes down to the decision that if he were told that he (as the A-body person) could choose who would have the post-operative torture, he would choose B, being convinced by his psychological angst arguments. My view is that, in Williams' "dump and restore" variant of the experiment, I would opt out because both A and B are dead, but in the "brain transplant" version, assuming I'm confident of surgical success, I'd choose the easy life for the body with my brain in it any day.

**Further Questions**
At Birkbeck, we are supplied with a Commentary that expatiates on aspects of the passage under discussion and asks (and sometimes sketches answers to) various questions. Because answering questions not surprisingly tests comprehension, I'll seek to answer some of the questions and you must let me know whether you agree with the answers :-

**Question**: Say which of the following seems the best account of what will have happened in the initial "third person" account of Williams' thought experiment:

(a) A turns into B, and B turns into A.
(b) A comes to inhabit B’s body, and B comes to inhabit A's body.
(c) A and B will both die and two other people emerge, say A+ and B+.

Can you think of any other descriptions that fit what happens more closely?

The fourth description (d) sought by the question would be that all sorts of horrid mental events happen to A and B, including total amnesia, character change and implantation of someone else's memories, but that A-body-person remains A and B-body-person remains B.

Description (a) is inadequate because what makes A or B a person includes both somatic and mental attributes, and the post-operative individuals are hybrids of A and B; the function "turns into" loses this sense of what's happened.

As discussed above, in the case of the "brain transplant", I would favour (b) as the best description, while for "brain copying" I would favour (c).

In case of brain transplantation it is not possible for any of A's conscious experience to remain behind in A's body, which rules out (d). The effect on A of finding himself in B's body, with different somatic, sensory and hormonal structures to get used to, would be exceedingly traumatic, but, though A would necessarily lose consciousness while the operation was performed, we (or at least I) can just about imagine him waking up feeling highly disorientated, but still feeling, as in (b) that he was A and not, as in (c), A+.

One reason I'm reluctant to support (b) in the "brain copying" example is the (science-fiction-) fact that multiple copies can be made of A's mental structures and parked in as many X_i-body-persons as we wish, and can they all be A? However, and I believe this to be a very important point, while we're making use of the technology, we could go further and use our matter-copier to clone A's physical brain multiple times and wire him up in the X_i-body-persons, or go the whole hog and clone the whole of A producing the set \{A^*_i\}. Would all these A^*_i’s feel themselves to be A? I think it's clear that, subject to the caveat in the next paragraph, they would, though, like twins, they would instantaneously diverge into separate persons as they developed further.
One caution - given how little consciousness as a phenomenon presumably derived from physical processes is understood - we should not get carried away by our thought experiments; for all we know we may be suggesting techniques that are not just impractical but impossible (as, for instance, would be the case if there was some immaterial soul that animates a person and that couldn't be cloned or copied by our ingenious devices). The metaphysical assumptions as to what makes up a human being are not made explicit in Williams' paper. It does seem, though, that the first-person perspective tacitly assumes that there's something more to "me" than the contents of my brain and that it's this "me" that continues in my original body as the (highly disturbed and utterly disorientated) person with a new mind. The first person perspective might almost be described as "exchanging minds", which is an even more difficult concept to get our heads round that "exchanging bodies".

**Question**: Williams considers different things that A and B might say if they were asked to choose the fate of the A- and B-body-persons, and also what they would say about their willingness to undergo the experimental operation. Summarize these various options in your own words.

Williams considers various responses (which I detailed above) all based on the "exchanging bodies" paradigm that A's consciousness ends up in B-body-person and vice-versa. From this third-person perspective, A chooses wisely if he chooses pleasant things to happen to B-body-person. Whether A would like to play this game would depend on what he thought of B's body and whether he found the prospect of occupying it appealing.

**Question**: If, when you answered the earlier exercise, you thought that this (ie. A and B changing bodies) was not the best description of the case, has the series of choices and estimates of outcome changed your mind? If not, what do you think might be wrong with the way these choices are described?

As Williams later points out at the end of his essay, the "exchanging bodies" paradigm is made more plausible by the perfect symmetry of the operation, whereas we might stop the procedure after we've copied B into A but before we've copied A into B. Also, though we've followed A & B's prudential thought-processes, we've not put ourselves firmly enough in their shoes to experience the angst they would feel at the possible results of the experiment for them, and hence may have missed out on the possible continuing A-ness of A-body-person.

**Question**: Williams considers ways in which one might resist this conclusion about the second of the perspectives on the example (ie. considered from the first-person perspective, at no stage does it seem as if you and the other change places). The arguments here are careful and sometimes dense, but you should read them over several times, and try to summarize them in your own words.

Williams doesn't seem very convinced that the first-person conclusion can be resisted, because, firstly, as A views the prospect before him, as summarised in Williams' (i) to (vi), at no stage does his fear abate, and in particular not at step (vi) which from this
perspective appears to be something happening to someone else. The key point is after (v) when B is very much alive and kicking, but A has not been reconstituted. Where is A, and who is A-body-person if B is still B? How do we resist the conclusion that A-body-person is still A? Williams doesn't hold out much hope for doing so, but recognises that someone committed to the exchanging bodies view will take it that step (vi) involves the re-introduction of A, who had at sometime prior thereto dropped out of the plot. Williams' problem is just when this might have been - he thinks it unlikely after (i) - amnesia - or (ii) - character change - and thinks it would have to be after (iii) - fictitious memory introduction - or (iv) introduction of memories modelled on B's. I've argued above that Williams underplays the catastrophic nature of these changes and that he ought to be more sympathetic to A dropping out of the picture after (i).

**Question:** Williams presents several arguments against the strategy based on thinking of our concept of a person as sometimes undecidable. Summarize his points in your own words.

This strikes me as the most difficult part of the essay. Williams rejects the idea that abandoning hope of clarity on the grounds of formal undecidability may provide some comfort. While it may supply some to third party observers of A's situation, it will supply none at all to A who is, potentially at least, intimately and disastrously involved. To clarify the situation, Williams asks us to consider a future situation S at which we (in fact, I) may or may not be involved. I will only feel fear at the prospect of S if I expect to be involved (and, of course, if I believe there is something in S for me to fear). If some dread event is going to befall one of a company of people, of whom I am one, my apprehension will rise to fear in proportion to my imagining that the one will be me. Moreover, I know that this situation will resolve itself and that it will either involve me or not. Williams also says that I may be neurotically apprehensive of some indeterminate ill from some range of possibilities, or even of some nameless horror, but the common factor for me to display fear at the prospect of one of these things happening is that they should happen to me. When I think that S may involve me, I am able to imaginatively project myself forward to my involvement in the event.

Williams now returns to our experiment, and finds A's predicament to differ from the indeterminacy in the examples just cited. It is not like the nameless horror, since that, whatever it was, was definitely going to befall me, nor is it like the probability case, where it would either involve me or not, depending on how the situation worked out. A's (my) fear of torture in the experiment seems neither appropriate (as it would be if I knew myself to be A-body-person) nor inappropriate (if I knew myself not to end up as the A-body-person), nor is it appropriate for me to be dispassionately equivocal, since the stakes are too high.

Williams seems to argue that if I try to imagine myself present at some situation S at which it is formally undecidable whether I will be there or not, then if my effort of what he calls projective imagination is successful, then I have effectively decided the situation in the affirmative, while if I am unsuccessful I've decided it negatively. Personally, I cannot see what is demonstrated by one's ability, or lack thereof, to work
oneself up into a lather of apprehension. I find a later argument even more confusing. Williams states that "material objects do occasionally undergo puzzling transformations which leave a conceptual shadow over their identity". An example of what this is supposed to mean would have helped. Whatever this is supposed to mean, we are to imagine that we are sentimentally attached to some object which undergoes this strange transformation, so that afterwards we are supposed now to feel ambivalent concern for it - neither as we did before, yet not totally disinterested. Not surprisingly, Williams does not think this a fair model of the situation envisaged when I'm not sure whether or not I'll be present at situation S, but that if I am something nasty is going to happen to me. I will not feel ambivalent concern for the person involved - I will feel terror on account of it possibly being me!

**Question**: Williams dismisses this second, conventionalist, way of evading the problem (of being precise in our definition of a person). What is his reason for dismissing it?

Williams' view is that whatever forensic decision is given by third parties on this subject, this is of no use to me who is vitally involved. If the conventionalists have decided that the best candidate for being A after (v) is A-body-person, to whom nothing further happens (apart from being tortured, of course!), the fact that there is a much better candidate after (vi), namely B-body-person, will be no consolation, as we had no necessity to proceed with step (vi) at all, so, from a forensic perspective, he might have been left as the unfortunate A-body-person.

**Question**: In the final paragraph of his essay, Williams tentatively suggests a resolution to the conflict between the two cases. Say what his conclusion is, and what argument he uses to reach it. Finally, give some reasons for either agreeing or disagreeing with him.

Williams' comes down tentatively on the side of the first-person perspective, and says that if he were A, he would choose for the torturing to happen to B-body-person. His reasons are that we were deceived by the symmetry of the third-person perspective experiment into considering it as an exchange of bodies, whereas if the experiment had been conducted asymmetrically (eg. only as far as the equivalent of (v)), we would not have been so convinced that this was the correct description of the procedure. As I have previously stated, I think Williams gives insufficient weight to the drastic changes that come over A- and B-body-persons, and that if the focus was on brain-transplantation, which is allegedly parallel to the experiment being performed, he would be on the side of the body-exchangers.

**Theo Todman**