

**Transplant Thought-Experiments:  
Two Costly Mistakes In Discounting Them**

[This is an un-edited and unformatted version of the paper that appears in the *South African Journal of  
Philosophy*, 33(2), pp 189-199.]

Simon Beck  
Philosophy Dept  
University of the Western Cape  
Private Bag X17  
Bellville, 7535  
Email: [sbeck@uwc.ac.za](mailto:sbeck@uwc.ac.za)

**Abstract**

‘Transplant’ thought-experiments, in which the cerebrum is moved from one body to another have featured in a number of recent discussions in the personal identity literature. Once taken as offering confirmation of some form of psychological continuity theory of identity, arguments from Marya Schechtman and Kathleen Wilkes have contended that this is not the case. Any such apparent support is due to a lack of detail in their description or a reliance on predictions that we are in no position to make. I argue that the case against them rests on two serious misunderstandings of the operation of thought-experiments, and that even if they do not ultimately support a psychological continuity theory, they do major damage to that theory’s opponents.

**Section 1: The Transplant**

‘Transplant’ thought-experiments, in which the cerebrum is moved from one body to another have featured in a number of recent discussions in the personal identity literature. Although they are different from the traditional ‘body-swap’ thought-experiments like the one with which Locke started the modern debate in that they envisage the part of the brain that supports distinctive psychology being transferred rather than just the psychology itself, they have usually been used in a similar role. That is, they are presented as offering confirmation to some form of psychological continuity theory of identity (PCT) which holds your continued existence as the person you are to be a matter of the persistence of your distinctive psychology,<sup>1</sup> rather than the persistence of the organism that is your body. But recent arguments have contended that this is not the case. Any such apparent support founders due to a lack of detail in their description, or a reliance on predictions that we are in no position to make. I wish to argue that the case against them is unsuccessful and even if they do not ultimately support a PCT, they stand to do major damage to that theory’s opponents.

---

<sup>1</sup> This is a very rough indication of what the PCT claims, but rough indications of the competing positions will suffice for the purposes of this paper. Where details are needed, as later with the ‘Person-Life View’, I will provide them. The details of the PCT can be found in Parfit (1995), Shoemaker (1984) and Beck (2011).

Locke's original idea involved us being asked to imagine the soul of a prince, with the prince's psychology, 'entering and informing' the body of a cobbler. According to Locke, 'every one sees' that the person in the cobbler body after the transaction would be the prince. He takes this to illustrate that notions of human identity and personal identity come apart and accordingly presents an account of personal identity as a matter of 'sameness of consciousness' (Locke 1975: 340). In 'The Self and the Future' (1970), Bernard Williams offered a technologically updated version of the story, eschewing souls as the mechanism for the transfer and instead asking us to imagine it being achieved by scanning the information stored in one brain and writing it into another. The result would be the same; the receiving body would have the psychology – memories, desires, beliefs, traits, projects and emotions of the person from whom the information was extracted. Williams's strategy is devious, however. Having presented the scenario as one that we would intuitively describe as a body-swap, he suggests we have been misled, and that another presentation leading to an opposed description is the one that should influence us. I will return to this alleged misleading later. These two versions of the body-swap occur in the debate between a psychological and bodily continuity account of identity, as does the similar presentation of Sydney Shoemaker. While they will be of some interest in my discussion, more pertinent is a scenario that may be even easier to imagine.

This occurs in the debate between the PCT and animalist accounts of identity – accounts like that of Eric Olson that take your persistence to be a matter of the continued existence of a living organism (Olson 1997). In these, the cerebrum is used as the mechanism of transferring psychology – removing it from one body and placing it into a recently decerebrated but otherwise viable body - that is, into a distinct living organism.

You have been diagnosed with cancer in an advanced state, and you have nothing to look forward to but a few months of intense, worsening pain followed by certain death. Nothing, that is, until a brilliant young surgeon offers to transplant your cerebrum, which is still free of cancer, into the head of an accident victim whose own cerebrum is damaged beyond repair. The operation is very safe, she assures you, and in all likelihood the result will be someone whose arms, legs, face and other parts are different from yours, but who has your cerebrum and, most importantly, your memories, character and other mental features. The rest of you, a brainless being that can still breathe, digest, and do whatever a human being can do without being conscious, will become the property of the local medical school, and will likely be used for experiments. (Olson 1997: 52)

The response that says you will survive is widespread – even Olson confesses to sharing the intuition which suggests that animalism is false. He thinks this is misplaced, though his particular reasons do not concern the detail provided in the description of the thought-experiment.<sup>2</sup> I want to turn to arguments that suggest it is particularly problematic.

## **Section 2: The need for detail – and why the needed detail might never be available**

Thought-experiments of this kind have been widely used, but they do sometimes face the resistance of those who claim to be unable to imagine what is being asked of them. Though Williams's case is ultimately against

---

<sup>2</sup> I discuss and respond to his reasons in Beck (2004).

any force that body-swap thought-experiments might have, he begins with some interesting considerations as to how they can be made more easily acceptable.

There are certain limitations, particularly with regard to character and mannerisms, to our ability to imagine such cases even in the most restricted sense of our being disposed to take the later performances of that body which was previously *A*'s as expressive of *B*'s character; if the previous *A* and *B* were extremely unlike one another both physically and psychologically, and if, say, in addition, they were of different sex, there might be grave difficulties in reading *B*'s dispositions in any possible performances of *A*'s body. Let us forget this, and for the present purpose just take *A* and *B* as being sufficiently alike (however alike that has to be) for the difficulty not to arise; after the experiment, persons familiar with *A* and *B* are just *overwhelmingly struck* by the *B*-ish character of the doings associated with what was previously *A*'s body, and conversely. Thus the feat of imagining an exchange of bodies is supposed possible in the most restricted sense. (Williams 1970: 161-2)

There are other considerations as well, of course, before we would respond that *A* and *B* have swapped bodies. Especially, Williams points out, we need assurance that the memories the *A*-body person has have actually come from *B*'s experiences and are not mere artificially generated memory-like experiences designed to match *B*'s earlier experiences (1970:162). Williams's brain state transfer device is envisaged as providing this assurance. These considerations are important, but I am more interested in the use Williams makes of the beliefs of the people in the society inhabited by *A* and *B*, and the idea that it is easier to imagine the person in the *A*-body as *B* if the people around that person think he is *B*.

Sydney Shoemaker uses the same factor in an even more explicit way in arguing for the PCT in his debate with Richard Swinburne. He adds the following details to the bare facts of the transfer of brain states.

We are to imagine that in this society going in for a body-change is as routine an occurrence as going to have one's teeth cleaned is in ours. It is taken for granted by everyone that the procedure is person-preserving. One frequently hears remarks like 'I can't meet you for lunch on Tuesday, because that is the day for my body-change; let's make it Wednesday instead'. All of the social practices of the society presuppose that the procedure is person-preserving. (Shoemaker 1984: 109)

Adding these details, he suggests, 'enhances the plausibility' of the view that a total brain-state transfer (BST) would indeed be person-preserving (Shoemaker 1984: 108).

Neutral parties to the debate, as well as Shoemaker's opponents, might well start becoming uneasy at this point. For the use of thought-experiments to count for anything in the debate, everyone seems agreed that they need to be phrased in terms that do not beg the question in favour of any particular theory. What would be unacceptable would be any stipulation of the identity of a particular individual in the imagined scenario, no matter how subtle. And it would not be that surprising if this charge was laid at the door of Shoemaker and perhaps Williams as well.

In Shoemaker's scenario, although we are not told explicitly that the person in the *A*-body after the procedure is *B*, we are told that everyone takes it for granted that the person is *B*, and that all of the institutions of society are grounded in this assumption. So, if we play the game and place ourselves in the society, we are likewise obliged to see them as *B* or we are not listening to the description provided. In that way, it might be argued, the question

of identity that is being asked is answered by stipulation. Although it is more subtle, it is no less infelicitous than Williams's stipulations in his second presentation in 'The Self and the Future', where he suggests that amnesia is produced *in A*, implying that it is still *A* after the procedure despite the loss of memory (Williams 1970: 172). But whether or not it is still *A* is what is meant to be up for grabs.

This argument moves a bit too quickly, however. Shoemaker's case is not quite as simple as the one envisaged here. He distinguishes between the way *we* use terms and the way *those in the narrative* use them, allowing that the two might come apart, and thus that we are not obliged by considering the narrative to accept their usage.

But, while we are not so obliged, he argues as follows.

There is no clear sense in which they can be said to be mistaken about a matter of fact in regarding the procedure as person-preserving. If we confronted such a society, there would, I think, be a very strong case for saying that what *they* mean by 'person' is such that the BST-procedure *is* person-preserving (using 'person' in *their* sense). ... But there would also be a strong reason for saying that what they mean by 'person' is what we mean by it; they call the same things persons, offer the same sorts of characterizations of what sorts of things persons are, and attach the same kinds of social consequences to judgements of personal identity... But if they are right in thinking that the BST-procedure is person preserving, and if they mean the same by 'person' as we do, then it seems that *we* ought to regard the BST-procedure as person-preserving. (Shoemaker 1984: 109-110)

That is, at least, not a simple stipulation of identity in favour of Shoemaker's psychological theory. But, while the distinction between our usage and their usage of terms is a fundamentally important one, it may not be enough to remove uneasiness that there is something illegitimate in the method of thought-experiment being stipulated here – something which affects the force of Shoemaker's argument. Kathy Wilkes's (1988) misgivings about the method in the context take hold at this point. She expressed concerns that are germane to the issue of whether or not his case is sound. There are reasons for believing that it is not clear that their usage is our usage, or, to stick more closely to the detail of her concerns, it is not clear what their or our usage would be. Shoemaker stipulates that they use their terms in a specific way, but that stipulation (according to Wilkes's reasoning) is one he is not in a position to make.

She suggests that any such envisaged counterfactual phenomenon as body-swapping would need a mass of information before the phenomenon could be 'established'. Her particular example is the case of a person splitting into two, but it is not difficult to extrapolate her concerns to the case of body-swapping and the Transplant.

It is obviously and essentially relevant *to the purposes of this thought-experiment* to know such things as: how often? Is it predictable? Or sometimes predictable and sometimes not, like dying? Can it be prevented? Just as obviously, the background society, against which we set the phenomenon is now mysterious. Does it have such institutions as marriage? How could that work? Or universities? It would be difficult, to say the least, if universities doubled in size every few days, or weeks, or years. Are pregnant women debarred from splitting? The *entire* background here is incomprehensible.  
(Wilkes 1988: 11)

She goes on to ask what seems a crucial question, 'When we ask what *we* would say if this happened, who, now, are 'we'?' Shoemaker's blithe assertion that 'all of their social practices' presuppose the procedure to be person-

preserving needs much more information to be provided before we can possibly recognize their usage as ours. Wilkes's implication is that we can never do so.

I think that there are two related assumptions about thought-experiments and how they work that play a crucial role in Wilkes's case. One is also to be seen in Williams's and Shoemaker's thought, as well as elsewhere in the literature. Both of them are mistaken, however, and this bears importantly on the significance of transplant thought-experiments.

### Section 3: Mistaking us for them

The first assumption is one that explicitly appears in Wilkes's writing. Here is how she describes how thought-experiments are used in investigating the plausibility of a philosophical theory.

It may be appropriate ... to ask a 'what if ...?' question. Such a question typically postulates an imaginary state of affairs, something that does not in fact happen in the real world. Put another way, in the modern jargon, we imagine a 'possible world' in which the state of affairs actually occurs – a world like our own in all relevant respects except for the existence in that world of the imagined phenomenon ... Then we try to draw out the implications – 'what we would say if' that imagined set-up were to obtain; that is, if we inhabited that possible world. (Wilkes 1988: 2)

The assumption comes out in the final sentence: that thought-experimenters are after what we would say *if we lived in the fictional world*. It is this that we are not able to discover since it would need knowledge of all the details of the kind she points to. We would need to know what it is to live with the new phenomenon and that is what we can never know from where we stand – thus she asks, 'Who, now, are we?' That is what she is indicating when she says that the entire background is incomprehensible. And that is why notions like splitting into two and swapping bodies cross the line from philosophy into fairy story and offer no support to any particular theory of personal identity.

The demand of knowing how to live with a phenomenon in order to know whether and how our concepts apply in its presence may be too hard to meet. It would certainly seem to require something more like a novel rather than the usual few-line vignette, and even then it is unclear if Wilkes would be satisfied. But Wilkes's dissatisfaction is misplaced, because the demand itself is misplaced. What we are after is *our* concept, not the concept that people in the society would have. We test the limits of our concept by whether or not it we are able to apply it when certain of the conditions in which it is usually applied change, but that does not mean that we are constrained in any way by how those in an imagined scenario apply their terms. Adding details to the scenario about how those in it use their terms might make it easier for someone who claims to be having trouble imagining the scenario to get around their difficulty – that is what Williams and Shoemaker see themselves as doing – but it is not a crucial factor in whether or not the implicit conditions underlying our application of a concept are met. Why, for instance, is it necessarily a problem for the possibility of a person splitting into two if universities and marriage would need to be handled differently if everyone were to do so? How does a

university's student population doubling in size if all of its current students go in for a split mean that we cannot understand the idea of one person doing so under very special circumstances? Wilkes says that these things are 'obviously and essentially' relevant to any splitting thought-experiment. But not only are they not obviously relevant, they are not relevant at all to one that is carefully and simply enough described.

There are other good reasons for keeping descriptions minimal. As Williams sensibly explains, the thought-experimenter is trying to prevent their respondents from being distracted by irrelevant factors. That the body into which your brain or psychology is transplanted might be very different from yours can prevent you from envisaging yourself in it. But that does not mean that *no* such transplant can be person-preserving. Nichols and Bruno point to research that indicates how easily details can distract us and make us susceptible to errors that we simply do not commit when the scenarios we are considering are kept abstract (Nichols and Bruno 2010: 20).

It may also be true that how we react to a scenario in applying our current concepts to it is not how we would actually apply them were we in the scenario. There is good evidence from social psychology to this effect. But that is neither here nor there when it comes to thought-experimenting.<sup>3</sup> Mastering a concept involves being able to apply it to (not necessarily *in*) counterfactual situations. When we are unable to do so – when the situation really is incomprehensible and we just don't know what to say – then we have gone beyond its limits. But that is for the respondent to say or to display; it is not for the critic to insist that someone is unable to apply their concepts when they do so consistently and with ease, as we do all the time when reading or watching science fiction.

#### **Section 4: Mistaking thought-experiments as confirmers**

The second assumption is, as I said, related to the first and some of these comments bear on it as well. It is an assumption that explains the demand for detail that is expressed by Wilkes, and her requirement that a thought-experimental phenomenon like splitting or body-swapping must be *established*. The assumption is that a thought-experiment functions as *evidence for* a philosophical theory. For a thought-experiment to play such a role, it seems obvious that the phenomenon it appeals to can only have authority if established in the way Wilkes demands, but sees as impossible in our instances. Once again, though, the assumption is misguided. This is not to say that some philosophers have not seen their thought-experiments as working in just this way – Locke, for instance, and Shoemaker can be seen as doing this. It is also how Williams views his first presentation, and his criticism of that presentation turns on the assumption. Let me explain this last point briefly as a way to approaching the issue of why the assumption is wrong.

Having presented his version of a psychology transfer, and going to great lengths to suggest that the only rational description of it is as a body-swap, Williams says that we should not be persuaded by the thought-

---

<sup>3</sup> As I argued in Beck (2006).

experiment to accept a psychological theory of identity. This is because our response to the scenario, and this resulting conclusion, has been artificially achieved.

The apparently decisive arguments of the first presentation, which suggested that *A* should identify himself with the *B*-body-person, turned on the extreme neatness of the situation in satisfying, if any could, the description of 'changing bodies'. But this neatness is basically artificial; it is the product of the will of the experimenter to produce a situation which would naturally elicit, with minimum hesitation, that description. By the sorts of methods he 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.

(Williams 1970: 179-80)

That is, even though we can easily apply our concept of same person, and naturally see the *A*-body person as *B*, there are other scenarios which might make us less happy about the theory that personal identity goes with psychology, like one in which we end up with two people who are psychologically continuous with *B*. The thought-experiment thus fails in its task of confirming the psychological theory.

Sure enough, the scenario does not prove the psychological view to be correct. But that is not something we should have ever expected it to do. What it is in a position to do is to provide indirect support to that view by showing that its opposition is in trouble. Williams's favoured view is that personal identity goes with continuity of body – what the scenario he painstakingly sets out shows is that our concepts do not require this to be the case. His theory makes bodily continuity a necessary condition of identity; the scenario makes clear that we do not have to see things this way, while still applying our concept. The thought-experiment works as a counter-example to that theory. That there are other scenarios that count against the rival psychological view does not affect this one being a counter-example at all. Since that is the whole case Williams has against the body-swap thought-experiment,<sup>4</sup> the experiment itself is left unscathed and Williams's theory is left in serious trouble.

My interests are more in the methodological point here, however. This is that the point of a thought-experiment of the kind that draws out intuitive responses is to illustrate that we don't *have to* see things in the way a particular theory says we do. It is not, or should not be, to show that we *must* see things in the way a particular theory suggests. Establishing that would be a much more difficult task, with perhaps the difficulties Wilkes foresees. But performing the task of offering a counter-example – as a refuter – to a claim of necessity has no such onerous requirements. It need only present conditions, as minimal as you like, in which we have a case in which we can apply a concept in the way the theory in question implies we can't. The artifice of the experimenter to which Williams takes exception is perfectly legitimate, given this end. And once again, the possibility of adding in further details that make the concept more difficult to apply – unless they are somehow implicit in the details already specified – does not affect the point that it can apply in conditions in which the theory in question implies it cannot.

---

<sup>4</sup> He famously has another presentation that is supposed to induce a different response from the reader, but this is the only direct point he offers as to the faultiness of the body-swap thought-experiment. I take on other aspects of his thought in Beck (1998).

In these ways, both Williams's and Wilkes's dismissal of transplant thought-experiments fail to succeed. I wish now to look at another attempt to blunt the force of a thought-experiment of this kind against a theory. The argument in which it occurs does not overtly state this as its aim, but like Williams's case it is used in support of a rival theory to the psychological one. And I will argue that its failure is related to the two mistaken assumptions I have just discussed.

### **Section 5: A new method, but old problems**

The argument in question is Marya Schechtman's in her book, *Staying Alive*. There she presents a view of personal identity, the 'Person Life View' (PLV) as a preferable alternative to the PCT. Part of the case is to set out a 'new method' for thought-experimentation (2014: 149-151). The precise details of the new method are not spelt out, though she does point to certain distinctive features of it, and argues that transplant thought-experiments either support her theory, or are harmless, when seen in light of the new method. I will argue in this section that she makes both of the mistakes I have highlighted. In the next, I will suggest that the transplant, without the mistakes, damages the PLV. Before doing so, it would be worth having some details concerning the PLV itself.

To be a person is to live a 'person life'; persons are individuated by individuating person lives; and the duration of a single person is determined by the duration of a single person life. (2014: 110)

The initial idea is that a person life is the sort of life lived by an enculturated human being. It follows a typical development from dependent infant, through the development of physical and psychological attributes which would at some stage include the attributes featuring in the PCT and which might also at some stage be lost. It accepts that humans can live very different sorts of lives, but points to a very general shared form of development (and in that way, even allows some non-humans as persons).

Importantly, the PLV also emphasises that a person life is lived in a culture and in interaction with other persons. Part of being a person is being engaged in characteristic interactions with other persons and against a background of social and cultural institutions; Schechtman talks of this as having a place in 'person-space' (2010: 279; 2014:114). This becomes significant in the dispute with the PCT. Although the existence of person-space requires that there be people with sophisticated psychological capacities, it includes other humans who do not share them, like someone in the late stages of senile dementia or a hydrocephalic child.

Schechtman suggests that *person life* should be seen as a cluster concept (as Chiong has characterised the concept of a biological life (Chiong 2005)). It is a cluster of biological, psychological and social functions which work together, but – unlike on the PCT or Olson's animalism – none is necessary and sufficient by itself for living a person life. While all three functions are usually coincident, they can come apart, and someone can still live a person life in the absence of any particular one.

Instead of assuming some one of these as the relation that constitutes our identity, we should think of identity as constituted instead by their interactions with one another. On the standard approach the fact that biological, psychological and social continuities are intertwined is seen as a complication which makes it difficult to determine which relation constitutes continuation. On the cluster concept model the integrated functioning *is* the true nature of the relation that constitutes the conditions of our continuation. The existence of the individual types of continuity in their ‘pure’ form is in fact a degenerate case of the more basic relation that contains all three. (Schechtman 2014: 150)

Her ‘new approach’ will focus on the interactions between these continuities, rather than on whether or not they can come apart. The picture presented so far is a slightly disingenuous description of the PLV, however. The details of her discussion suggest that social continuities are *always* required for the continuation of a person life, while either of the other two may be missing. The requirement is two of the continuities, one of which is that the individual is able to ‘occupy the same place in person-space’ in that they are treated by others as the same person. As she says, both the method and the PLV emerge better in considering cases like the Transplant.

She presents a case based on Olson’s experiment, but far more detailed than any we have met so far.

Sometime in the future an environmental toxin reaches levels at which it begins to regularly cause liver failure in a large segment of the population. A technique is developed to clone healthy livers from an individual’s own tissue, transplant techniques are improved, and liver transplants become common. Later the toxin begins to attack other organs, and these are regularly cloned and transplanted as well. Eventually, it attacks all tissue but the cerebrum (which is somehow protected). Fortunately, cloning technology has developed to the point where healthy whole organisms can regularly be cloned. Moreover, the development of clones can be accelerated and directed so that the result is an adult human body that looks almost exactly like the individual from whom the genetic material was taken, but which lacks a cerebrum. The cerebrum of the diseased individual is then placed into the cerebrum-less skull of the cloned human, carrying with it the individual’s beliefs, values, desires, memories and so on. This operation inevitably and immediately leads to the end of the biological life of the organism that used to contain the cerebrum. Everyone refers to this operation as a ‘full body transplant’ and sees it as the limiting case of the transplantation of individual organs. Just as it is assumed that a person survives when she gets a new liver or kidney or heart, it is assumed that a person survives when she gets a new body (or, strictly speaking, cerebrum-complement). After post-surgical recovery the patient typically returns to her family, friends, job and hobbies. (Schechtman 2014: 151-2)

There are echoes of Shoemaker’s technique in that we are told how everyone in the situation calls it a body-swap. (I have defended inclusion of those details, though not Shoemaker’s reasoning about them.) And the details provide at least a little of what Wilkes demanded.

Schechtman’s response to her own scenario is not, as per the usual way and as reflected by Shoemaker, to say that we would say that someone would survive this procedure – that it is person-preserving. Rather, what she says is that *the PLV has* the implication that it is person-preserving (2014: 152). She points out that the intuition that the person who emerges from the operation is the ‘natural target of the questions, concerns and relationships’ that applied to the original person is ‘nearly universal’ (2014: 152). That is, everyone agrees that the required social continuity is present in the scenario. This is her practice throughout the discussion – she never asks, ‘What would we say?’ but only describes what the PLV says and discusses the plausibility of that.

She does admit that the plausibility of the PLV’s implications concerning some scenarios may be problematic. While that may sound like an approach towards the account I have given of thought-experiments as refuters, it is

nothing like that at all. That is because the idea that we might not agree with the implications of the PLV concerning a scenario has very little, if any, significance. Our judgements concerning survival in these cases are ‘provisional’ (2014: 154) – such a judgement is ‘only a prediction’ (154) which may well be very wrong. In the cases in which the PLV is troubled, she suggests this is particularly likely. It may still have some explaining to do (and she goes to some lengths to do that), but given the unsatisfactoriness of the evidence against, any damage will be very limited.

Why are our judgements of survival to be discounted? The reasons will be familiar, and there is a clue to them in the previous paragraph – they are predictions based on insufficient evidence. What we are doing in making these judgements is predicting how people will be treated in circumstances crucially different from our current ones, and this is something that we cannot know with any degree of reliability.

It might be, for instance, that unforeseen pressures or developments would make it impossible to really treat transplant products of the sort described as if they were the original person.  
(Schechtman 2014: 154)

All of this is familiar from the discussion of Wilkes. Whether or not our concept applies depends on the way those in the situation envisaged would actually apply theirs. That is the assumption Wilkes makes and which I argued was mistaken – we are investigating our concept as it stands and providing a theory of that notion of personal identity or as close to it as can be provided, not a theory of a notion that others might hold or that we would come to hold if things were different. The mistaken step is explicitly taken by Schechtman. In demanding that the story be detailed enough to include a convincing background as to how the technology enabling the transplant developed, she writes

The point of this back story is obviously to make it plausible in the minds of those who inhabit the world depicted (*and so, as readers of the story, in ours*) that the transfer of the cerebrum to a decerebrated skull will look merely like the extension of other kinds of transplants of a sort that already exist and are clearly person-preserving.  
(Schechtman 2014: 152, my italics)

But as I have argued, we are perfectly capable of applying our concepts to circumstances in which the characters involved apply theirs differently. So whether or not their view is plausible to us does not prevent us from applying our concept to their situation. And whether or not we would actually be like them when things changed is a sociological or psychological issue and not the conceptual one that we face in this debate.

The second assumption that I criticised is also to be seen in Schechtman’s discussion. The signs are clear in her demand for detail – the passage just quoted can be seen as requiring, like Wilkes, that the counterfactual phenomenon be established for it to count. Other points in her discussion reinforce this, especially that her ‘new method’ is about illustrating the way in which the various aspects of survival interact rather than considering whether they may come apart – it is about providing a form of confirmation of a theory, in this case the PLV. ‘The real question,’ she says, ‘is not so much “would the person survive in the case described?” but rather “is the case described coherent and plausible?”’ (Schechtman 2014: 153). The case, that is, in which the PLV says that they would survive.

But, as I argued, thought-experiments can do very little by way of confirming a theory. That you can tell a plausible story that is consistent with a theory does little or nothing for it other than illustrate it. Schechtman is not unaware of this – she accepts that once it is stipulated that the person who survives the procedure is treated as the original, ‘the implication that the cerebrum donor survives as the whole body recipient follows immediately’ (2014: 153), though she still seems to think if we can ‘truly envision’ things happening as described this does something for the theory (2014: 153). But even if a thought-experiment can do little directly for a theory, I have pointed out that it might well do damage to other theories. Before she sets out her new method for thought-experiment, Schechtman seems to acknowledge this. Transplant thought-experiments, among other things, show the need for Korsgaard and Olson to radically alter their theories to make room for a ‘forensic unit’ – a locus of responsibility (2014: 58-65). But the new method and its reliance on the two assumptions and their related reasoning takes the sting out of that.

If our judgements about survival in a counterfactual scenario are no more than not well-founded predictions, and are provisional on whether or not they are the ones people in the situation would actually make – something we can have little confidence about – then they can hardly be taken as defeating a theory. The fact that we all agree intuitively that the person who emerges from the transplant is the cerebrum donor means very little, and should hardly bother the animalist. The metaphysical issues that animalists then raise against the PCT (and which affect the PLV equally) become much more significant as favouring that theory since we have no serious reason to question it in the first place. But I go along with Schechtman’s first thought – that the scenario does great damage to animalism. Our ready response that the cerebrum donor survives (perhaps made easier by the described attitudes of those in the story), when the animalist holds they do not survive, means that the animalist does not capture anything close to our notion of personal identity in their theory. Schechtman may be willing to pay the price of losing this advantage in order to make her theory proof against thought-experiments that threaten it. But losing the advantage is a very high price to pay and is a price paid for the wrong reasons. Her denial of the refuting force of the thought-experiment is based on erroneous assumptions about the significance of the views of the fictional characters and the role of thought-experiments in relation to our concepts.

### **Section 6: The Transplant takes another victim**

Let us take a brief look at the sort of thought-experiment which Schechtman suggests might be more difficult for the PLV. This is the case in which, unlike the simple case presented above, the body of the cerebrum-donor remains alive in a vegetative state in hospital. There are two possible background stories we could provide, according to Schechtman. In the one, the cerebrum recipient takes up the thread of the original life and is treated by everyone as the original person while the individual in a vegetative state is viewed as a husk of the original, subject to very few person-related concerns. This is slightly problematic for the PLV since in the case of actual advanced dementia, an individual in exactly similar vegetative state is seen by the PLV as being the original person, and not just a husk of them. In the second, the vegetative individual is subject to many more person-related concerns; as a result, the PLV implies that there are now two originals – we have a form of fission. Schechtman’s response is first to say that it is very difficult to predict ‘just how products of this kind of

operation would be treated or perceived' (2014: 155). But she thinks that both eventualities can be envisaged, subject to a story being told, and makes suggestions as to how the PLV can cope. In the first it would have to acknowledge some, but not damaging, conventionalism in whether or not someone survives (2014: 156ff). In the second, she takes a leaf out of Parfit's book and suggests that the original does not survive, but that not surviving is not like normal death (2014: 158ff).

Her discussion acknowledges that the PLV has to do some work in order to cope with this case, but it does not have great urgency, given the speculative nature of the troublesome evidence, which is never seen as actually being in a position to show the theory to be wrong. Other opponents might want to suggest the coping mechanisms are less than satisfactory, but I want to focus rather on another scenario that she does not consider.

Consider a society in which a cerebrum transplant operation occurs. The cerebrum of a person, with their fully developed psychology is transplanted into another body, leaving her original body as a living organism with whatever support it requires to function in a minimal way as before. This society sees this organism as the original person and treats her accordingly, just as they treat someone who has lost her capacities and is in very advanced dementia. They ignore the troublesome individual who keeps turning up at the hospital entreating them to take notice of her. They are firm in the belief that that they are acting correctly and eventually resort to a restraining order against this annoyance. According to the PLV there is no problem, since there is only one individual who takes up the original person life and who is the subject of the required social continuity and has one of the other two continuities, the biological one.

But, of course, there is a problem. It is obvious, from where I stand (and, I suspect, from where you stand as well), that they are doing something monstrous albeit with a firm conviction otherwise. They have got a matter of fact about personal identity wrong. Since the PLV implies that this is not so there is something wrong with it as a theory of personal identity, or rather, it is not *our* theory of personal identity. It lands in the same camp as animalism, even if it has come closer to getting things right than that theory does. Schechtman would have to say that we cannot truly envision this society; we cannot plausibly see it as developing from ours. I think that it should be clearer by now that that would just be a sulk. A bit of imagination helped by the power of religion gets us there. But I also think that is beside the point. We can understand very well what is going on in the scenario, described in the little detail provided. Our concept of personal identity fits easily into place. If there is something relevant and deeply impossible that is being glossed over, that would need to be shown, and I can't see what it could be.

## **Section 7: Concluding**

I've considered two prominent attacks on the significance of Transplant thought-experiments in the current debate. Both of them founder because of misplaced assumptions about thought-experiments and how they might work. Once those assumptions are put aside, we can see how they might bother certain participants in the

debate. They don't show, as some have thought, that the PCT is correct. But they can do crippling damage to its opponents - animalism and the Person life View.

## References

- Beck, S. 1998. 'Back to the self and the future', *South African Journal of Philosophy*, 17(3), 211-225.
- Beck, S. 2004. 'Our identity, responsibility and biology', *Philosophical Papers*, Special Edition, October 2004, 3-14.
- Beck, S. 2006. 'These bizarre fictions: thought-experiments, our psychology and our selves', *Philosophical Papers*, 35(1), pp29-54.
- Beck, S. 2011. 'Causal Co-Personality: In Defence of the Psychological Continuity Theory', *South African Journal of Philosophy* 30(2): 244-255.
- Chiong, W. 2005. 'Brain death without definitions', *The Hastings Center Report*, 35(6), 20-30.
- Locke, J. [1694] 1975. *An Essay Concerning Human Understanding*, ed. P. Nidditch, Oxford: Oxford University Press.
- Nichols, S. and Bruno, M. 2010. 'Intuitions about personal identity: an empirical study', *Philosophical Psychology*, 23(3), 293-312
- Olson, E. 1987. *The Human Animal*. New York: Oxford University Press.
- Parfit, D. 1995. 'The unimportance of identity'. In H. Harris (ed.), *Identity*. Oxford: Clarendon Press.
- Schechtman, M. 2010. 'Personhood and the practical', *Journal of Theoretical Medical Bioethics*, 31, 271-283.
- Schechtman, M. 2014. *Staying Alive*. Oxford: Oxford University Press.
- Shoemaker, S. 1984. 'A Materialist View'. In Shoemaker, S. and Swinburne, R. *Personal Identity*. Oxford: Blackwell.
- Wilkes, K. 1988. *Real People*. Oxford: Clarendon Press.
- Williams, B. 1970. 'The self and the future', *Philosophical Review*, 79(2), 161-180.