

Note: This is an un-formatted version of the article that appears in *South African Journal of Philosophy*, 35(2), 2016, 132-140.

## **Reconsidering a transplant: A response to Wagner**

**Simon Beck**

Philosophy Department, University of the Western Cape, Bellville  
South Africa

Author email: sbeck@uwc.ac.za

Nils-Frederic Wagner takes issue with my argument that influential critics of “transplant” thought experiments make two cardinal mistakes. He responds that the mistakes I identify are not mistakes at all. The mistakes are rather on my part, in that I have not taken into account the conceptual genesis of personhood, that my view of thought experiments is idiosyncratic and possibly self-defeating, and in that I have ignored important empirical evidence about the relationship between brains and minds. I argue that my case still stands and that transplant thought experiments can do damage to rivals of a psychological continuity theory of personal identity like Marya Schechtman’s Person Life View.

### **Some background**

Nils-Frederic Wagner (2016) takes issue with my argument (Beck 2014) that influential critics of ‘transplant’ thought experiments make two cardinal mistakes. He responds that the mistakes I identify are not mistakes at all. The mistakes are rather on my part, in that I have not taken into account the conceptual genesis of personhood, that my view of thought experiments is idiosyncratic and possibly self-defeating, and in that I have ignored important empirical evidence about the relationship between brains and minds. Once all of these are taken into account, the case that I have suggested in support of these thought experiments and a psychological continuity theory (PCT) of personal identity disappears, and what emerges is a case for its rivals—especially the one I highlighted for criticism, Marya Schechtman’s Person Life View.

Transplant thought experiments are those in which the cerebrum of one individual is envisaged as being successfully transplanted into the “decerebrated” body of another. They have often been taken as central to the case for showing the PCT to be correct, in that we respond that we would go along with our cerebrum and psychology into a new body. This intuitive response is consistent with how that theory explains identity—as a matter of overlapping psychological connections forming the continuity that marks personal persistence. My focus on such thought experiments was not because I see them as crucial

to supporting the theory of personal identity that I think comes closest to getting things right. Rather, it was because they seemed to be most suited to meeting the demands of critics of thought experiments (and that theory) in the personal identity debate. In fact, the ones on which I focused were ones proposed by those critics themselves. When everything else is said, these are ones they are prepared to consider. I am persuaded that Locke's old story of the prince and the cobbler reveals important things about the concept of personal identity,<sup>1</sup> but others have told me that they could not grasp what Locke was asking them to imagine. So I focused on thought experiments that appeared much easier to imagine. Since they were the products of the critics' imagination, I hoped to put that complaint aside.

They were of even more interest as it seemed to me that, once some methodological points were understood, they could do damage to the theories favoured by some of those critics.

### **The "Two Mistakes"**

The mistakes I identified were in the thought of Kathleen Wilkes, Bernard Williams and (more recently) Marya Schechtman. Wilkes focuses on fission thought experiments rather than transplant-type ones, but her arguments can be easily applied to the latter as well. She objects that they require us to make predictions that we are in no position to make. The thought experiments posit worlds utterly unlike our own and then require us to predict how our concepts would operate ("what we would say") if we existed under those conditions. To do that, though, we would need to know all sorts of things that we cannot possibly know:

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).

That sounds at first like a reasonable response, but it seemed to me to include some strange thinking. The point of thought experiments that ask what we would say is to find out what we ultimately believe about something—the implicit conditions underlying our application of a concept. That is, they attempt to test our deep commitments—which roughly correspond to the necessary conditions for our application of a concept. Many conditions will be met in such an application, but we are more committed to some than to others; we find

---

<sup>1</sup> As a result, I am not at all sure that I am committed to the 'brain cause' view of psychological continuity that Wagner says I insist upon (20). At the same time, I am very much a materialist and dispute only a few of the things he has to say about the relationship between brain, body and distinctive psychology.

out which by separating usually co-instantiated conditions. That is what counterfactual scenarios are meant to achieve: we are faced with a scenario in which some usual condition no longer obtains and asked to consider whether our concept still applies.

But if that is so, we are not being asked for a prediction about what concept people would have if conditions were to change. That is, we are not being asked what we would say if we were to exist under those conditions—how our concept would change if things changed radically. We are being asked whether we are able to apply our concept as we understand it *now* to a situation where a particular condition no longer obtains. This is not something impossible to predict, unlike the answer to Wilkes's question. We either respond readily, or we do not, just as when we read science fiction. I called Wilkes's emphasis on what we would say *in* the situation (rather than what we say *about* it) the error of mistaking us for them.

The second error I referred to as that of mistaking thought experiments as confirmers. Williams (1970, 179–180) rejects body-swap thought experiments like Locke's<sup>2</sup> on the grounds that there are other possible outcomes similar to that envisaged which create difficulties for a psychological continuity theory of identity, and which the thought experimenter suppresses. I argued that this strategy only has any teeth if thought experiments are understood as (somehow conclusively) showing a theory to be true. If there are thought experiments that create difficulties for the PCT, then a body-swap thought experiment with which it is consistent does not show it to be correct. But the point of the body-swap case to which Williams objects is to show that having the same body is not a necessary condition for being the same person. This consequence still stands even if there are other arguments against the PCT and so Williams's case misses the mark. And I suggested that the idea that thought experiments could show a theory to be true is simply implausible: telling a story that is consistent with a theory only illustrates it—it provides no more confirmation for that theory than for any other theory consistent with the story. Where thought experiments have their teeth is in showing theories to be false—as being inconsistent with our conceptual scheme. The damage caused by Williams's body-swap thought experiment to his own theory that bodily continuity is necessary for identity stands.

I suggested that Wilkes' and others' insistence that thought experimental scenarios are usually hopelessly under-specified rests on a similar mistake. As we have seen, she thinks that once the detail required for us to respond adequately to them is supplied, they become incomprehensible. The detail is needed to “establish the phenomenon” to which they appeal. But establishing the phenomenon in detail is only required if the experiment is to provide evidence *for* a theory (otherwise detail is only required, for heuristic purposes, to assist those

---

<sup>2</sup> He does not reject brain-transplant thought experiments, only those that use another mechanism like a brain-state transfer device to get the psychology from one body to another.

people who cannot understand what they are being asked to imagine<sup>3</sup>). To show it false, we only need an instance where we apply our concepts in ways that the theories say we cannot. As long as no deep impossibilities are being hidden by a thin description, then no elaborate detail is needed to achieve this end.

### **Is the first mistake really a mistake?**

Wagner denies that either of these is actually a methodological mistake. As to the first one, he contends that we do indeed need to know all of the social details that Wilkes demands if we are to get to what we think about persons and their identity. What I have got wrong is to assume that the concept of *person* exists in a vacuum and to take this “to assume that personhood would not be a different concept if it were shaped by vastly unlike circumstances” (21). I have adopted a view that there are certain intrinsic person-constitutive features that hold across all possible worlds, ignoring the importance of the relational and social nature of persons and how personhood is embedded in particular social practices. That I ultimately draw the conclusion that personal identity is a matter of intrinsic (psychological) features means that I am actually begging the question (22).

The details Wilkes demands, he suggests, are all-important:

If the social ontology of personhood was vastly different in a possible world, then there is little to gain for an investigation of “our” concept of a person. And so, universities and marriages and all the other socially constructed observer-dependent facts...work in a specific way precisely because they are so closely tied to our social ontological concept of personhood. We cannot treat these observer-dependent facts as though they were observer independent, facts holding across all possible worlds. Marriages and universities in a world in which splitting is commonplace would potentially be unlike ours in a great many different ways because the underlying concept of a person would be significantly different. Perhaps in such a world your split sister was also married to your husband and your split brother was also allowed to take your upcoming metaphysics midterm because you were seen as one and the same person. The reason, then, for why we need to hear more about the practices that people in such a possible world would have is not to know what ‘they’ would think, but because we cannot figure out what ‘we’ think without information about the social functionality of the transformed beings (Wagner 2016, 21-22).

One aspect of this response is immediately puzzling. This is the charge that I assume that the concept *person* exists in a vacuum and is the same across all possible worlds. It is puzzling since that is something I explicitly deny. My claim that it is a mistake to confuse us with them rests on the suggestion that their concept might be different from ours, given the different circumstances with which they have to cope. Especially in circumstances radically different from ours, beliefs and practices which relate to how a concept is applied are very likely to be different. And I acknowledged that we have no adequate way of working out how

---

<sup>3</sup> Just enough detail until they get it.

those practices will change, and so are in no position to speak with any authority<sup>4</sup> about what their concept would be after radical change. But my point was that their concept is not what thought experiments are meant to investigate. Thought experiments are after our concept and its commitments. And we do not need to have special insight into what people in another world would think to find out what we think. To have it spelt out for us what they in the scenario think might make it easier for us to apply our concept—as I discussed in detail—but it is not a requirement for getting to our conceptual commitments.

Even if we did need to know what they would think, Wagner’s contentions do not affect a central strand of my case that there is a mistake going on here. This is that Wilkes (and Wagner and, to a lesser extent, Schechtman) base their misgivings on our inability to predict how our concept would be affected by *radically* different circumstances. The problematic thought experiments at which they gesture involve individuals who split all the time—universities that double in size every few days, spontaneous unpredictable splitting, or your split sister being also married to your split husband. These are all very complex (though not really all that difficult to understand, and ones that may not really lead to different notions of personhood). But that complexity is simply beside the point. A thought experiment of the kind in question tries to isolate specific conditions—in our case whether being the same person requires being the same organism. That needs a case in which we agree that this is the same person, but it is not the same organism. The splitting need not be spontaneous, or multiple, or involve any change to social institutions or be unpredictable in any way—the thought experiment can specify (or describe in detail for the troubled) simple “social functionality”. Because of that, the practical difficulties grounded in complexity that all of these critics raise are simply not relevant. Schechtman takes the same line as Wagner, however:

If we encountered a single instance of doubling we undoubtedly would not know how to react, and if it became common we would become different sorts of beings with an entirely different (and to me not-yet-imaginable) social organization and another way of life...Such beings might come to be, but they will probably not be persons (Schechtman 2014, 164).

But this is hyperbole, and an illustration of my complaint. We can understand a single case perfectly well—we react to (the story of) a transplant quite equably. That we might actually be troubled if it were to happen does not reflect on what we think now about such a case. There is no reason to have to imagine a wildly complicated world in which the phenomenon occurs, just as there is no reason why we *have* to think about splitting being common. And the leap to splitters probably not being persons is simply out of order. The reasons being

---

<sup>4</sup> That is, we do not know how people will actually behave should circumstances change radically. But that does not mean we cannot stipulate what they think for the purposes of a thought experiment scenario. We do have authority in that sense, and that is enough for getting at what we think now.

suggested here for not taking the relevant thought experiments seriously are not close to being adequate to the task.

### **Is the second mistake really a mistake?**

Wagner also takes issue with what I called the mistake of seeing thought experiments as confirmers. Their role as I outlined it is to undermine strong claims made by particular theories. He suggests I present no positive argument for this view, that it is a highly idiosyncratic one, and that it can even be seen as self-defeating.

Let me begin with the charge of idiosyncrasy. According to Wagner, this is because my view of thought experiments as refuters would only apply if theories of personal identity were making claims of necessity. This, he suggests is not the case—not all of them are committed to necessity claims (Wagner 2016, 23). I only erroneously think they do because of my assumption that only intrinsic features matter to personal identity. I have denied that I make that assumption.<sup>5</sup> That aside, the theories that I suggested in my paper were affected by the thought experiments under discussion—those of Williams, Olson and Schechtman—all *do* make necessity claims. Williams thinks that continuity of body is necessary for identity; Olson thinks the same of being a continuing human organism. Schechtman's theory has a very different feel to it, but even she is offering, in her Person Life View (PLV), a theory that she stresses is a metaphysical one and which makes claims very like necessity claims. The PLV is the view that to be the same person is to continue to live the same person life. *Person life* is a cluster concept, comprising psychological, biological and social continuities which usually function together and no one of which is itself necessary for survival if the other two are in place. But that does not mean she is not making necessity claims—two of them *must* be in place. In fact she actually does require one of them in all cases (and perhaps in all worlds). She writes,

*it is essential* to the judgment that a person survives a “whole-body transplant” that the transplant product is able to pick up the thread of the life of the person who enters surgery. This can happen only if the transplant product is accorded the appropriate place in person-space; that is, if she is treated as...the continuation of the original locus of concern (Schechtman 2014, 152, my italics).

Her view on how thought experiments work is different from mine, as I described in the original paper, but this does not mean she is avoiding claims of necessity. In her view,

---

<sup>5</sup> Although it is very widely held to be a fundamental principle that identity must be intrinsic—cf. Noonan (1989, 152 and 164), Williams (1960, 45—at least according to Parfit 1984, 267), and early Schechtman (1996, 31).

once the case stipulates that the person who results from the transfer is treated as and responds as the original person the implication that the cerebrum donor survives as the whole body recipient follows immediately (Schechtman 2014, 153). If that implication follows immediately, then we are dealing with something very like a claim of necessity—like enough to necessity to make my account relevant.

What might be unusual about my theory is that others think that thought experiments are in the business of *establishing* necessity claims. Jonathan Ichikawa and Benjamin Jarvis write, “On our view, intuitions [in response to thought experiments] are judgments of necessity, and we see no in principle objection to the idea that we can know them a priori” (Ichikawa and Jarvis 2009, 223). George Bealer thinks that they “present themselves as necessary” (Bealer 1998, 207). But a single case can never establish necessity. That does not need any argument. And it does not seem to me that the claim that thought experiments do not offer significant positive confirmation of a theory needs much more. A single experiment at most offers a little confirmation of a theory in other fields. It also confirms every other theory consistent with the experimental result. A story offers less confirmation than that, and also confirms to the same degree every other theory consistent with the story. Thought experiments only become interesting to the argument once they threaten a theory, and that is exactly my point that they function as refuters.<sup>6</sup>

Wagner sees my view as a denial of any positive role for thought experiments, despite my insistence that that they offer indirect support for a theory by undermining rivals. All theories of identity, he says, have faced fanciful cases that have indicated problems in them. “This, however, has been done in order to support a rival view by showing that this view does not suffer from problems in the scenario” (Wagner 2016, 23). That is exactly my view on what they can do: it goes without saying that for it to offer indirect support for your theory, your theory must be consistent with the thought experiment. To propose a thought experiment which destroyed your own theory in your efforts to counter your opponent would be a most rash strategy indeed.

Wagner suggests my view of thought experiments being primarily refuters makes them pointless: “I am unaware of any theory that has not been haunted by dubious counter examples” (23). But this argument turns on the thought experiments involved being dubious ones, and that all problems are of equal weight. That will not work as a reply here—the thought experiments on which I focused are the critics’ own ones, raising problems that they acknowledge need solving.<sup>7</sup>

---

<sup>6</sup> I am not the only person to hold this view. Roy Sorenson calls them “alethic refuters” (1992, 135).

<sup>7</sup> Olson goes to great lengths to reject the force of the transplant on the grounds that reasoning from it to the falsity of his view requires reliance on an inference from moral responsibility to identity that holders of the PCT cannot make because of their commitment to fission arguments (1997, Ch 3); to no avail, as I argue in Beck (2004).

The charge of my method being self-defeating is as follows. A thought experiment can count against a theory and indirectly in support of another only if the second theory is better able to deal with the scenario. “But then the thought experiment serves to confirm the theory, and Beck does not want to have it that way” (Wagner 2016, 23). But that is exactly how Beck does want to have it and how he has set the method out—it is just that that confirmation does not offer much positive support, unless the confirmed theory were the only one that could possibly deal with the scenario. That might be the case in some instance, but would need much further argument<sup>8</sup> and remains in line with my account of the method as turning on refutation. Thought experiments only become interesting once they reveal a problem.

Wagner ends his case here by reading me as implying that transplant cases support the PCT by counting against all theories other than the PCT (23). But that bears no relation to any argument of mine. I only suggested that such cases count against two theories, and I think that there are many considerations other than these cases that count in favour of the PCT.

### **The implausibility and impossibility of my transplant**

I detected both of my “mistakes” occurring in Schechtman’s use of a transplant thought experiment, and then presented a development of her transplant case that I argued counted against her Person Life View (while the PCT could cope happily with it). Hers is a detailed story:

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, 151–152).

---

<sup>8</sup> Just as Locke has to go on to argue that his prince and cobbler case does not provide equal confirmation to identity being a matter of continued immaterial substance as to his own theory. He needs further thought experiments to achieve this.



Her PLV yields the judgement that, since the recipient of the cerebrum has both psychological and social continuities, they continue the person life and so survive, despite the lack of biological continuity. She acknowledges that there may be a problem for her view concerning the donor organism which continues to exist, since it is in an analogous situation to a patient with advanced dementia—who the PLV would also count as continuing the person life and so surviving. She has a response to this, as I described. But I put forward a different possible development which I thought was more difficult to cope with.

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 (Beck 2014, 198).

In line with my account of thought experiments, I contended that the onus would be on the critic to show that the development I had introduced to Schechtman's thought experiment concealed some deep problem, or otherwise the PLV's counterintuitive verdict would count against the theory. Wagner takes up the challenge: "The seeming force against the 'Person Life View' stems from a conceptual implausibility and an empirical impossibility" (25).

Before looking at what Wagner believes these deep problems are, a point I made at the outset should be revisited. This was that the thought experiments that formed the focus of my paper were not *my* ones, but ones outlined by those critical of the theory I favour and aspects of its methods. Rather than take on the particular development of a transplant case that I have outlined, Wagner argues against the assumptions of transplant thought experiments in general. That is perfectly acceptable as a philosophical move outside of this dialectical context, but hardly fair as a response to my particular paper. When he sets out the "false assumptions" that I make he is taking on assumptions that those to whom I was responding had granted me by making them themselves. They amount to common cause. In the context of the argument, what is required is isolation of impossibilities *only I* and not the other users of transplant thought experiments have assumed. Nothing meeting that description is ultimately to be found in his response.

That being said, I do not agree that any of us are guilty of the charges laid, or that our scenario is as confused as Wagner would have it.

The first mistaken assumption of my transplant that he identifies is that I grant the cerebrum traditional features of the soul. "What is presupposed...is a conception of the

psychological subject as a discrete and unified object located within the brain and as such (at least in principle) removable from the rest of the body” (Wagner 2016, 25). What this presupposition ignores is how cognition depends on far more of the body than just the brain, even though the brain is pivotal in maintaining your psychology. He outlines how Schechtman in an earlier paper (1997) pointed to a “Distributed View” where the brain is seen in cognition as the heart is seen in circulation—as the central organ at the core of a system distributed throughout the entire organism. There is thus no clear-cut distinction to be made between the brain and body as the transplant requires (25).

Empirical evidence backs up how important bodily features are to how we navigate the world and how they impact our distinctive psychology. Our immune systems distinguish between our and other tissue, and rejection has been observed in brain tissue transplants. This means we cannot expect psychology to be transplanted with a cerebrum. What I have done wrong is to submit to a wishful “impulse to impose conceptions of the mind formed within the context of dualism onto a materialist ontology” (Wagner 2016, 25).

In dismissing this latter impulse, Wagner appears to have forgotten his own views about the importance of the social ontology in which the conceptual genesis of personhood takes place. The genesis of our concept of personhood is very much in a context where dualism is a possibility. To rule out, as conceptually impossible, disembodied minds is to choose to investigate a notion of mind and personhood that is by no means a common one. Nor is that common understanding of mind obviously at odds with a materialist ontology: the task of the materialist is to locate the concept of mind with all of its quirks—or as many of them as can be consistently retained—in a physical world. It is worth noting that arch-materialist David Armstrong sets as a requirement on any theory of mind that it allows the possibility of disembodied existence—precisely because that is part of the concept that he is trying to locate in the material world (Armstrong 1993, 19–20). Not all of that inheritance may be redeemable: there are aspects we may have to give up as having no place in reality or as conceptually mistaken—just as the PCT gives up the principle that identity is always what matters.

The empirical evidence that Wagner points to is important, but none of it marks the transplant scenario as impossible. Our cognition relies (in actuality) on all sorts of relationships between brain and body. As he says, our distinctive psychologies and way of making our way through the world are shaped by the bodies we have and the interaction between brain and the rest of the body. But, although Wagner is suggesting that my psychology is the result of the particular body I have, the evidence presented requires no more than a particular *type* of body—one that is very like mine. To think like I do, I need the

central organ<sup>9</sup> that governs my thought—my cerebrum—and a body that has the same sort of relation to my brain as mine does. But that is what transplant thought experiments presume: Williams made and resolved all of these points back in “The Self and the Future”, and thought experiments nowadays use twins and exactly similar bodies of new matter as run-of-the-mill features. Schechtman’s version uses a cloned body. Such a body rules out the tissue rejection objection as beside the point (and it was never a conceptual impossibility, as Wagner acknowledges (26).

Is it not important that “Beck and Schechtman do not share the same presuppositions when it comes to the cerebrum transplant thought experiment that Beck takes to cause trouble for the ‘Person Life View’” (Wagner 2016, 26)? It is not important, because it is not true. The Schechtman who rejects the transplant is the Schechtman of 1997, who (interestingly) accepts the assumption attributed by Wagner to me that personal identity must be a matter of intrinsic features (a point she suggested then that the PCT cannot include) (cf. Schechtman 1996, 31). The Schechtman to whom I am responding is that of 2014 who explicitly shares my presuppositions about the transplant scenario—it is *her* scenario after all, one she uses to argue for (among other things) the superiority of her PLV over animalism. The PLV suggests that you *do* survive the transplant, as we have seen—as long as the survivor takes up your place in person-space. Schechtman and I share a view on “the closely intertwined relation between brain and body” where Wagner sees a contrast (25)—she does not (despite his contention) see a relation between brain and a particular body as “crucial in preserving identity” (Wagner 2016, 25). 2014 Schechtman differs with me on how many teeth thought experiments have, but not on the presuppositions or even the outcome of the transplant. As a result, I cannot see any actual grounds for Wagner’s claim that her PLV has an advantage over the PCT. Rather, relying only on her presuppositions, a different version of the transplant like that I set out in my paper can be described, in response to which the PLV yields a strongly counterintuitive judgment. Unlike the PCT.

There is one point that Wagner raises that is specific to my version of the transplant. For argument’s sake, he grants that a cerebrum transplant might be successful as envisaged and be accompanied by the social circumstances I specified, then goes on to argue that the thought experiment would not be useful to my case.

But even if it did happen, one could argue that it is not obvious that those people would be wrong in claiming that the original person would be the donor and not the recipient. What we would have discovered is not that that is how people in this other world think of things, but rather that is how “we” think of things. Take an example: you might say, “If I win the lottery I would quit my job in an instant” and then find out, when you do win the lottery that you actually do not want to quit. Similarly we might want to say that the recipient of the cerebrum would surely be the same person, but if

---

<sup>9</sup> Or, perhaps, just an organ *like* mine with the right causal history.

the experiment were performed it might turn out that we actually feel differently (Wagner 2016, 26).

However, we do not discover what we think in this way—we discover rather what people *would think* if it were to happen. I am not disputing that changing circumstances can change judgements. But that means that we discover what *they* think, or how thinking would change if circumstances were to change. That is exactly my point about distinguishing between us and them. We are after our concept—represented in how *we think* we would behave (“what we want to say” in Wagner’s terms) and not after theirs—which might be revealed in how those in the circumstances actually behave.<sup>10</sup> We may well not be able to predict with any accuracy how we would behave if circumstances changed drastically. Our intuitions have only representational authority—they reveal at best what we think; they are not direct routes to reality and do not have that sort of epistemic authority.<sup>11</sup> But it is only representational authority on which thought experiments rely, and all that they need to be effective refuters of a theory of personal identity, and indirect supporters of another.

## References

- Armstrong, D. 1993. *A Materialist Theory of the Mind*. Revised edition. London: Routledge.
- Bealer, G. 1998. “Intuition and the autonomy of philosophy.” In: *Rethinking Intuition*, edited by M. DePaul and W. Ramsey, 201–239. Lanham: Rowman & Littlefield.
- Beck, S. 2004. “Our identity, responsibility and biology.” *Philosophical Papers* (special issue): 3–14.
- Beck, S. 2006. “These bizarre fictions: Thought experiments, our psychology and our selves.” *Philosophical Papers* 35(1): 29–54. doi:10.1080/05568640609485171.
- Beck, S. 2014. “Transplant thought experiments: Two costly mistakes in discounting them.” *South African Journal of Philosophy* 33(2): 189–99. doi:10.1080/02580136.2014.923685.
- Ichikawa, J., and B. Jarvis. 2009. “Thought experiment intuitions and truth in fiction.” *Philosophical Studies* 142(2): 221–46. doi:10.1007/s11098-007-9184-y.
- Noonan, H. 1989. *Personal Identity*. London: Routledge. doi:10.4324/9780203428351.
- Schechtman, M. 1996. *The Constitution of Selves*. Ithaca: Cornell University Press.
- Schechtman, M. 1997. “The brain-body problem.” *Philosophical Psychology* 10(2): 149–64. doi:10.1080/09515089708573212.
- Schechtman, M. 2014. *Staying Alive*. Oxford: Oxford University Press. doi:10.1093/acprof:oso/9780199684878.001.0001.
- Sorenson, R. 1992. *Thought Experiments*. Oxford: Oxford University Press.
- Wagner, N.-F. 2016. “Transplanting brains?” *South African Journal of Philosophy*, 35(1), 18–27. doi: 10.1080/02580136.2015.1106705.

---

<sup>10</sup> I say *might* be revealed, since actual behaviour is not always the best guide to values (survival values or otherwise). The Milgram experiments reveal that people are willing to torture others when required to do so by an authority figure. That does not mean that they think it is right to torture others simply because someone in a white coat asks you to do so.

<sup>11</sup> This is the thrust of my account of how thought experiments operate and of how they can do so effectively despite our tendencies to self-deception in Beck (2006).

Wilkes, K. 1988. *Real People*. Oxford: Clarendon Press.

Williams, B. 1960. "Bodily continuity and personal identity." *Analysis* 21(2): 43–48. doi: 10.1093/analys/21.2.43.

Williams, B. 1970. "The self and the future." *Philosophical Review* 79(2): 161–80. doi: 10.2307/2183946.