

Martha Nussbaum and the Foundations of Ethics: Identity, Morality and Thought-Experiments

Simon Beck

School of Philosophy and Ethics
University of KwaZulu-Natal
Private Bag X01, Scottsville, 3209
South Africa
Email: beck@ukzn.ac.za

Abstract

Martha Nussbaum has argued in support of the view (supposedly that of Aristotle) that we can, through thought-experiments involving personal identity, find an objective foundation for moral thought without having to appeal to any authority independent of morality. I compare the thought-experiment from Plato's *Philebus* that she presents as an example to other thought-experiments involving identity in the literature and argue that this reveals a tension between the sources of authority which Nussbaum invokes for her thought-experiment. I also argue that each of her sources of authority presents further difficulties for her project. Finally, I argue that it is not clear that her thought-experiment is one that actually involves identity in any crucial way. As a result, the case she offers does not offer any satisfactory support for her view on the relation between identity, morality and thought-experiments, but we do gain some insights into what that relation really is along the way.

Introduction

While the debate around personal identity has long been of central interest to philosophers, its emphasis has started to show changes. This is reflected in recent major publications. One notable example is a focus on the relation between theories of personal identity and morality, as in David Shoemaker's *Personal identity and ethics* (2009). Another is the interest in the role of thought-experiments both in metaphysics and ethics, as in Kwame Anthony Appiah's *Experiments in Ethics* (2008). The seeds of these changes have, however, been around for quite some time, and both of the issues appear together in an important essay of Martha Nussbaum's published in the mid 90s. The context of that essay was a dispute (between Nussbaum and Bernard Williams) on how to understand Aristotle's attempt to base ethics in human nature. My paper is not intended as Aristotle scholarship and my concern is not really with how Aristotle should be read nor about who wins this dispute. My interest is that a central part of Nussbaum's argument draws together the independently important matters mentioned above – the nature of personal identity, how this relates to ethical issues and the methods we should employ to work all this out. I want to argue that by comparing and contrasting what Nussbaum says on these matters to what some others have said, we can learn a fair amount about them, and this may well offer us some enlightenment as to the merits of her case and on the issues themselves.

Although I will not pursue this aspect of the discussion, others might find further interest here from the point of view of the importance which Nussbaum's "capabilities approach" has attained in the fields of policy and development. The foundations for ethics indicated in our discussion might well be the metaphysical foundations Nussbaum envisages for that approach.¹

Section 1

Nussbaum's foundational endeavour and her response to Williams centres on the use of a certain type of thought-experiment in moral discussion. The thought-experiments as outlined by Nussbaum lead to conclusions about moral principles or positions based on considerations about personal identity. The arguments get their authority from the "broad acceptability across time and place" (Nussbaum 1995: 121) of these considerations and, more importantly, from the "self-validating" structure of the thought-experiments themselves. In the context of her dispute with Williams, Nussbaum appeals to this source of authority as providing the foundational force for ethical conclusions. In this way, she sees Aristotle as founding ethics in human nature. But she is firm in insisting that this is not to find a foundation in the science of human beings as Williams characterises Aristotle's aim (Williams 1985) because, she holds, the relevant considerations concerning personal identity are also matters of moral judgement, not independent scientific fact.

While I have indicated that my focus is not on who wins the Williams-Nussbaum debate on Aristotle, all of the other points raised in the synopsis just provided will play a role in my discussion. The central and clearest example of the kind of thought-experiment that will occupy our attention is the one Nussbaum identifies in Plato's *Philebus*.

The thought-experiment is Socrates' case against Protarchus's view that pleasure is the ultimate good for human beings. The dialogue runs as follows (*Philebus*, 21a-d). Nussbaum draws attention to a number of features of the argument that she deems crucial to it. First, that ethical work is going on. The conclusion of the argument is that pleasure cannot be *the* good. Her second point is that the argument relies on considerations of personal continuity:

Protarchus has been asked whether he would choose *for himself* a certain sort of life; he is asked to imagine *himself* living it. It now develops that Protarchus really believes (though ... he did not realise it) that this life is really not one that can be lived by a human being at all. But if this life cannot be lived by a human being, it cannot be the object of Protarchus' own self-wishing and self-choosing. The good life for him (for us) must first of all be a life for him (for us). (1995: 100)

Third, Socrates' technique draws out Protarchus's deep beliefs: "Protarchus has been brought to see that he really has believed all along that any life that could be a life for him must involve certain uses of practical reason" (1995: 100). The technique relies solely on the basis of Protarchus's own beliefs about himself and his life. Nussbaum's final point is that the argument works not just on Protarchus's beliefs, but on his *evaluative* beliefs – what in the personal identity debate are often called "survival values". She writes,

¹ I will not be pursuing this aspect since my conclusions about Nussbaum's case will be mainly negative, and thus offering no help to those seeking a metaphysical basis for the capabilities approach.

The life described seems too impoverished to be human because of what he cares about and thinks to be good, because he cares about these things so much that anyone who lacks them could not, he thinks, be identical with him, and would have to rank as a lower form of life. (1995: 100)

Nussbaum's stress on this final point is to lay the foundation for her response to Williams that Aristotle, when his turn comes, does not appeal to some external authority independent of morality in order to give it foundation. Instead, part of the authority of this thought-experiment comes from the broad acceptability and basic nature of the beliefs about his continued existence that Protarchus is revealed to hold deep down. The rest comes from the special structure of the case the thought-experiment presents (1995: 101).

The structure of the argument, Nussbaum contends, renders it self-validating. By entering into the debate with Socrates – by “accepting Socratic procedures” – Protarchus is already committed to the value of practical reasoning. He would then be inconsistent if he were to deny the conclusion that the good life requires respect for reason as well as pleasure.² Socrates' argument is a form of *reductio ad absurdum* of Protarchus' professed belief that pleasure is the good. So it is not simply Protarchus's subjective outlook that gives weight to Socrates' conclusion – all who do not opt out of the “form of life that the Socratic procedure exemplifies” are bound to accept it as well. In this way, according to Nussbaum, this sort of argument can provide a satisfactory foundation for ethical thought.

Section 2

That gives us a more detailed idea as to Nussbaum's view on how thought-experiments concerning identity can relate to moral discussion, and fills out what her claim that “certain ethical proposals go against our deep beliefs about identity” (1995: 105) is all about. Her view on the relation between thought-experiments, identity and morality contrasts with some other current views, and there is good reason for taking issue with her on a number of points.

Links between thought-experiments, morality and identity emerge right at the start of the modern discussion of issues of personal identity. John Locke strikes a chord with Nussbaum on the matter of claims of identity or non-identity involving moral judgements when he characterises the idea of personal identity as a “forensic” notion (Locke 1975: 50). His discussion of the conditions of personal identity (as opposed to the identity conditions for other types of thing) constantly places matters of responsibility alongside matters of identity, suggesting that he sees a conceptual connection between the two. And while his focus is on identity rather than morality, his choice of argument form strikes another chord – he favours thought-experiment.

... should the soul of a prince, carrying with it the consciousness of the prince's past life, enter and inform the body of a cobbler, as soon as deserted by his own soul, every one sees he would be the same person with the prince, accountable

2 For the sake of the argument I will let this go, but I am not convinced that Nussbaum is correct that there is an inconsistency here (not even a “pragmatic” one as Susan Hurley suggests (1993: 271)). For surely Protarchus can consistently use reason to show that something favours us *not* reasoning (for example, that we would be happier not reasoning)? That is no more problematic than arguing, in Latin, that something favours not using Latin.

only for the prince's actions: but who would say it was the same man? (Locke 1975: 44)

While this thought-experiment does not involve any change from human to some other kind of being, it is certainly meant – like Nussbaum's *Philebus* example – to evoke deep beliefs about the conditions of continuing existence. Locke thinks, as the quote suggests, that it shows that personal identity goes with continued “consciousness” rather than bodily continuity. That is to go a bit too quickly, however. What it shows is that someone who shares Locke's intuitions is more deeply committed to some form of psychological continuity rather than bodily or physical continuity as far as their existence is concerned. To put it in terms like those Nussbaum uses, and taking the experiment to apply to them to mirror the case of Protarchus, they do not see the life of the individual in the body continuous with theirs as their life because they care more about psychological than physical continuity. As with Protarchus, our subject may be unaware that they had this belief all along (we don't meet body-swappers every day, after all).

Like Nussbaum, Locke is of the opinion that these revealed deep beliefs about identity are of a broad enough appeal (“every one sees”) to lend them authority. This makes the analogy appear very strong between the two thought-experiments, and offers some explanation for the strong conclusion Locke draws and of which I was wary above. But my wariness was not without reason. It is worth stressing that all that the thought-experiment is in a position to reveal is *what somebody believes* – what their (perhaps unacknowledged) commitments are. This inevitably leads us to the point that others may have other commitments.³ Although in the case of Locke's thought-experiment these are rare, at least one writer claims exactly that.⁴ There is room for genuine disagreement; people may have different intuitive responses, even if (as in this case) there happens to be broad agreement.

What would such differences mean? They mean (barring avoidable misunderstandings of the description of the thought-experiment) that we have a case of different concepts of personal identity and personhood. Does that mean that our concept is right and theirs wrong? No, of course not. But things need not end in this relativist impasse. Both sets of respondents imply that there are certain entities – persons that exist over time – that meet the criteria matching their concept. There may or may not be such entities – we have, so far, only ideas of them. We need to see if these entities can be located in the world. The position is completely familiar from the failure of the Ontological Argument for God's existence. Some such entities can be located in more familiar and better-evidenced features of the world, others must be eliminated. This does not mean that each competing concept is not on to, or missing out on, some important feature of the world. But if it is, we need to go beyond the beliefs of our respondents and get independent support for the view embedded in this conceptual picture – and that is exactly the move that Nussbaum resists in her dispute with Williams.

Section 3

“But hold on,” I hear you say, “This is all very well. But doesn't Nussbaum also get the foundational authority in question from *the structure* of *Philebus*-like thought-experiments?” As reported above, she emphasises that they are *self-validating*. Since we

³ Nussbaum herself seems to acknowledge this when she writes, “Yet they can be ... denied, especially on account of theoretical prejudices or political interests” (1995: 102).

⁴ I am referring to Peter Unger. See Unger (1990) p.159 note 7.

are faced with a *reductio* of Protarchus' initial position, shouldn't we rather look to that source of authority? I will argue that the problems are not that easily avoided, but in the meantime let us take a look at another influential thought-experiment in the area that shares features highlighted by Nussbaum, and which is in a position to aid our investigation of the self-validating aspect of the case.

Derek Parfit uses a thought-experiment that also works as a *reductio*, also concerns personal identity and also leads to a conclusion that is moral in nature. He calls the thought-experiment, "My Division" (Parfit 1995: 42; 1984: 254).

Parfit assumes that he could survive with only half of his brain, justifying this on the grounds that people who have suffered strokes and thus lost the use of half of their brains can survive with very little psychological change. He also assumes that each half of his brain is able to support ordinary psychological functioning. On these assumptions, he points out, if half of his brain were destroyed and the remaining half transplanted into a body like his own we would have no reason to say that he did not survive: "in this *Single Case*, the resulting person would be me" (1995: 42).

He then outlines *My Division*:

Both halves of my brain would be successfully transplanted, into different bodies that are just like mine. Two people would wake up, each of whom has half of my brain, and is, both physically and psychologically, just like me. (Parfit 1995: 42)

Faced with this scenario, Parfit asks whether or not it would preserve "what matters in survival". He argues that we have to say it would.

In the *Single Case*, the one resulting person would be me. The relation between me now and that future person is just an instance of the relation between me now and myself tomorrow. So that relation would contain what matters. In the *Double Case*, my relation to that person would be just the same. So this relation must still contain what matters. Nothing is missing. (Parfit 1995: 42)

I said at the start of the section that *My Division* functions as a *reductio ad absurdum*. That might not be obvious from what has been quoted so far, but that it is indeed the case becomes clear when we consider what he has outlined in relation to the issue of identity, and when we consider that Parfit's target is the "Non-reductionist" about identity – someone who believes (amongst other things) that *identity* is always what matters. In describing the result of *My Division*, Parfit points out that he cannot say that "each of them is me". That is because they are two distinct people, and two things cannot be one. It also cannot be that one of them is me, because there is nothing that makes one, rather than the other, me. That would make identity an arbitrary matter. The only other option logically available would be to say that neither resulting person is me. But there is no difference between the resulting person in the *Single Case* who *is* me and either one of the resulting people here who *is not* me. The Non-reductionist, committed to the importance of identity, has to provide one of these answers – but none of them are acceptable. The result is that we have a *reductio*: "So identity cannot be what matters" (1995: 42).

In this argument we are not simply asked for intuitive responses that are taken to reveal our survival values. The point of Parfit's thought-experiment is not to draw forth an answer to the question of whether you would rather continue as you are or split into

two. The point is rather to illustrate some fundamental tensions - tensions between the assumption that identity is what is at issue and some fundamental (logical and other) principles. On the assumption that identity is what is at issue, we are obliged to offer one of the four descriptions of what has happened to him that Parfit outlines. The problem is that to say he is both survivors conflicts with the principle of the necessity of identity (one thing cannot be two); to say he is one rather than the other conflicts with the principle that identity cannot be an arbitrary matter; to say he is neither conflicts with the principle of the intrinsicness of identity – whether or not that is you cannot depend on whether or not *someone else* exists. These principles are not subjective survival values.

My point is that *My Division* is not an intuition-pump for producing your survival values. It does not turn on the question, “what would we say?” but instead illustrates that *whatever we say* we are in trouble. In that way, it works as a *reductio ad absurdum* of the “natural” view (Parfit 1984: 215) that identity is the crucial issue in the debate.

You may still be puzzled as to the relevance of all this, even if the similarities of argument structure are clear. But here come the insights that it offers for our concerns. If a thought-experiment and its accompanying argument have the structure of a *reductio*, then that argument will always work like Parfit’s – that is, independently of any reliance on the contingent survival values we might hold. There just is no role for those to play in the argument. The same goes for the case of Protarchus. If that argument is indeed a self-validating one, then what Protarchus believes deep down about his identity is quite irrelevant to its conclusion.

There is a more important point, though, than simply that Nussbaum has thrown in a red herring. Parfit’s “My Division” forces us to revise our beliefs – it takes a commonsense principle and reveals why we cannot hold it. It is not just that it does not reveal deep beliefs – it can only show a belief to be *wrong*. That’s the business of a *reductio*. So it’s not just that the Philebus argument does not rely on what Protarchus believes, it could not possibly reveal what he truly believes or is committed to – not, at least, in so far as it is self-validating. At best it can show that you cannot hold that pleasure is the only good *and* be prepared to reason about it.

At times, Nussbaum appears to acknowledge this. She characterises the conclusions of the various arguments she discusses as being that “certain ethical principles go against our deep beliefs about identity” (105), and that is what a *reductio* might lead to. She also acknowledges that one might opt out of Socrates’ form of life. But that nevertheless leaves questions about the whole project of founding ethics by this route. The route is inevitably a revisionist one – requiring the giving up of beliefs in the light of others, not the cementing of those other beliefs. And if all we can learn is what we cannot believe if we believe some other things, how will we ever get to any sort of foundation for ethics?

Nussbaum wants two parts of the Aristotelian project in one go: to “save the appearances” in retaining our commonsense views and to provide a foundation for ethics. That explains the various points she highlights, but this route does not allow both.

Section 4

Contrasting Nussbaum’s account of the *Philebus* thought-experiment with others in the literature has led to some interesting results so far. I want to look at a third contrast

now, with thought-experiments and related discussion in the work of Eric Olson. This will move us, initially at least, away from self-validating thought-experiments.

Once again, on the surface there is a great deal of consonance between Olson and Nussbaum. Olson fits in to something of an Aristotelian tradition in that his metaphysics is an overtly substance-based one, and in his playing down of any distinction between humans and persons. He also agrees with Nussbaum that the judgements operating in the context of thought-experiments about identity are moral ones. But this does not mean that there is anything like general agreement between the two. Olson insists that what the exploiters of intuitive responses to thought-experiments characterise as judgements of identity are in fact moral judgements that have nothing to do with identity at all, and serve only to mislead the discussion. He uses this example that he calls "The Transplant".

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)

Olson acknowledges that the broad intuitive response to this thought-experiment is that you would survive in the body of the cerebrum recipient. Indeed, that is his own immediate intuitive response. Nevertheless, he insists, you would actually survive as the brainless object of experimentation. The only reasons you are tempted to think otherwise are first, that you might confuse this case in your considerations with one in which a whole brain is transferred (in which case you would survive as the recipient, the presence of the brainstem indicating which organism is which). The second reason is the more important one. This is that you intuitively feel prudential concern for the cerebrum recipient that you do not feel for the brainless being, and you intuitively believe that the recipient (and not the brainless being) will bear responsibility for your actions. These intuitions involve what are obviously moral judgements, and are standardly taken to reveal deep beliefs about identity – to reveal that you really think that the recipient will be *you*. Olson wants to argue that this is to make an unacceptable jump, since identity simply does not follow from rational prudential concern nor from moral responsibility.

Nussbaum seems to be committed to the inference that Olson questions. The argument from The Transplant and the judgements of identity that he criticises there are of a very similar kind to those Nussbaum identifies in the *Philebus*. Because Protarchus does not value the jellyfish life enough, it could not be his. That involves what amounts to the same link between identity and morality in judging that because you do not value the brainless life, but do care greatly about the recipient life, the former cannot be yours. In this respect, then, Nussbaum and Olson take up opposing positions. This may not appear all that important, but it does have significance for our discussion

which is worth noticing. Given her acceptance of this link between identity judgements and moral judgements, and her professed approval of belief-revealing thought-experiments, it is very hard to see how Nussbaum can avoid committing herself to the standard response to The Transplant. And that gives her a *prima facie* commitment to a distinction between humans and persons: a distinction she is keen to avoid in her paper – especially as she wants to defend an Aristotelian foundation for ethics in human nature that is somehow a matter of moral judgement. Olson is busy trying to shed himself of precisely this commitment.

On the other hand, we still have not seen any reason as to why Olson holds that identity judgements and moral judgements come apart. The argument for that is one familiar from the previous section – My Division. Olson points out that My Division requires us to revise some of our basic moral principles, such as “You are responsible for only your actions”. Since Parfit’s fission offshoots are still (intuitively) responsible for his actions, but are not identical to him, the responsibility-identity link is broken and the standard argument from The Transplant falls through.⁵ The same goes for prudential concern – My Division shows how you can rationally feel prudential concern for someone who is not you. That you do feel such concern for the brain recipient does not imply that that person is you.

Now, Olson is aiming his argument at Parfitians – those committed to a psychological account of our identity. We may want to come to Nussbaum’s defence by pointing out that only they are, and she is not, committed to the conclusions of Parfit’s thought-experiments. Thus, perhaps she could maintain the responsibility-identity link by rejecting the argument of My Division. What makes this an unlikely way out is that (as we can see from the recent discussion of her views on the interconnectedness of morality and identity) she obviously shares many of the ideas that supporters of a psychological account cling to and take as distinctive of their position, and that she obviously approves of arguments of the form and substance of My Division. They are, after all, the very stuff of her case against Williams. At the least, she has a lot of difficult explaining to do.

There is a final point that comes indirectly from considering Olson’s views on identity and morality. As described, Olson argues that the judgements playing the crucial role in a thought-experiment purportedly about identity are only moral ones and not identity ones at all. I think we can ask whether this is not true of the *Philebus* thought-experiment. Let us go back to Nussbaum’s setting-up of the point.

In all of this, the role played by considerations of personal continuity and species membership seems to be as follows. Protarchus has been asked whether he would choose *for himself* a certain sort of life; he is asked to imagine *himself* living it. It now develops that Protarchus really believes (though ... he did not realise it) that this life is really not one that can be lived by a human being at all. But if this life cannot be lived by a human being, it cannot be the object of Protarchus’ own self-wishing and self-choosing. The good life for him (for us) must first of all be a life for him (for us). (1995: 100)

What makes us think that this is about personal continuity and not simply about moral preferences? Nussbaum’s italics indicate that the questions concern Protarchus. That it is Protarchus *himself* who must choose only suggests that they are *Protarchus’*

⁵ I think that Olson gets more out of this argument than he is entitled to and I argue that out in Beck (2004). But he has got enough right to offer the comparison that I need with Nussbaum.

preferences that we are talking about, not his identity. That he is to imagine himself in the situation adds nothing to that; to get at your moral preferences I will ask, “Would you do A if such-and-such were the case?” I am not asking about your survival values, only about your plain old values. That a life is not worth choosing because it lacks features that humans in general find important (“is really not one that can be lived by a human being at all”) takes us no closer to issues of Protarchus’ personal identity. A life of constant humiliation, totally lacking in dignity, is not a life worth choosing; but that tells us something about human values, not necessarily about personal identity and certainly not about human identity. Nussbaum is reading in what is not there –that is confirmed by Socrates’ final question: “Could we consider such a life desirable?” That is all that is really at stake in this case.

Conclusion

Nussbaum’s discussion of the Philebus thought-experiment is intended to endorse the view (supposedly that of Aristotle) that we can, through thought-experiments involving identity, find an objective foundation for moral thought without having to appeal to any authority independent of morality. By comparing her exemplar thought-experiment to other thought-experiments involving identity, I have highlighted a tension between the sources of authority Nussbaum invokes. A thought-experiment cannot work by revealing our deep commitments if it is, at the same time, self-validating. I have also argued that thought-experiments that work by revealing deep commitments need independent authority if their conclusions are to count as authoritative. On the other hand, those that are self-validating will be ones that question our commitments rather than ones that can show our commitments to be foundational. Neither kind stands to offer the support that Nussbaum envisages from thought-experiment. Finally, I have argued that it is not clear that her thought-experiment is one that involves identity in any crucial way. All things considered, then, the Philebus thought experiment does not offer any satisfactory support for her view on the relation between identity, morality and thought-experiments.⁶

References

- Appiah, K.A. 2008. *Experiments in Ethics*. Cambridge, MA.: Harvard University Press.
- Beck, S. 2004. “Our identity, responsibility and biology”. *Philosophical Papers*, Special edition, November 2004, 3-14. Available at <http://www.ukzn.ac.za/phil/OurIdentityResponsibilityandBiology.pdf>
- Hurley, S. 1993. “Martha Nussbaum: Non-Relative Virtues”. In Sen, A. and Nussbaum, N. (eds.), *Quality of Life*. Oxford: Oxford University Press, 270-276.
- Locke, J. 1975. “Of identity and diversity” in J Perry (ed), *Personal Identity*. Berkeley: University of California Press.
- Nussbaum, M. 1995. “Aristotle, Nature and Ethics” in J Altham and R Harrison (eds), *World, Mind and Ethics*. Cambridge: Cambridge University Press, 86-131.
- Olson, E. 1997. *The Human Animal*. Oxford: Oxford University Press.
- Parfit, D. 1984. *Reasons and Persons*. Oxford: Clarendon Press.

⁶ I am grateful to two anonymous referees for comments on an earlier draft of this paper.

- Parfit, D. 1995. "The Unimportance of Identity" in H Harris (ed.), *Identity*. Oxford: Clarendon Press, 13-45.
- Plato. 1975. *Philebus*. Translated by JCB Gosling. Oxford: Clarendon Press.
- Shoemaker, D. 2009. *Personal Identity and Ethics: A Brief Introduction*. Broadview Press.
- Unger, P. 1990. *Identity, Consciousness and Value*. New York: Oxford University Press.
- Williams, B. 1985. *Ethics and the Limits of Philosophy*. Cambridge, MA.: Harvard University Press.