

The Non-Identity Problem and the Admissibility of Outlandish Thought Experiments in Applied Philosophy

ADRIAN WALSH

University of New England, Armidale, Australia

University of Gothenburg, Gothenburg, Sweden

The non-identity problem, which is much discussed in bioethics, metaphysics and environmental ethics, is usually examined by philosophers because of the difficulties it raises for our understanding of possible harms done to present human agents. In this article, instead of attempting to solve the non-identical problem, I explore an entirely different feature of the problem, namely the implications it has for the admissibility of outlandish or bizarre thought experiments. I argue that in order to sustain the claim that later born selves cannot be harmed (since they are in fact different persons), one must rule inadmissible certain kinds of modally bizarre imaginary cases. In this paper I explore how one might justify such a constraint on outlandish cases and, in so doing, develop the outline of a model for distinguishing between admissible and inadmissible imaginary cases in philosophical debate.

Keywords: Thought experiments, non-identity, philosophical methodology, moral luck.

1. *Introduction*

The non-identity problem directs our attention to the obligations we might have towards people: (i) whose existence we cause in some relevant sense; (ii) whose circumstances, while tolerable, are less than ideal but (iii) who would not exist if those less than ideal circumstances were improved. In such cases determining whether or not a harm has been done to a particular person must be constrained by considerations of whether or not less harmful alternative courses of action would have led to that person existing. Parfit's description of the problem in *Reasons and Persons* begins with the assertion that the issue of which fu-

ture people will exist is dependent (at least in part) on when exactly the procreation takes place (Parfit 1984: 358). Parfit claims that in assessing the *relative harm* of the circumstances of any birth, one cannot compare them with other scenarios involving alternative policies and actions in which one would not have been born at all.

Much of the subsequent discussion has focussed on either the question of what the *proper object* of moral concern might be or the implications it might have for our obligations to future generations. Should it be states of affairs or individual persons? However, there is another significant feature of this line of reasoning that such discussions ignore; namely regarding the set of admissible scenarios for moral assessment. The debate tacitly assumes that *modal harms*—that is, counterfactual harms that are either nomically impossible or practically infeasible in the circumstances—are not relevant to the assessment of individual welfare. The refusal to countenance the outlandish or morally bizarre, I shall refer to as the ‘*nommic constraint*’.¹ While Parfit directs our attention away from worries about individual welfare, in this paper the critical focus will be on one of the methodological assumptions that drive arguments for shifting away from individual welfare.

The distinctive feature of the discussion herein is that it treats the non-identity problem as requiring, amongst other things, constraints upon the relevant thought experiments and imaginary scenarios one might employ. The outlandish would appear to be ruled out. I suggest that reconsidering the non-identity problem in terms of the limits it places on what counterfactuals are relevant, sheds fresh light on our understanding of the proper role of thought experiments in our moral reasoning and, in particular, the role of outlandish or bizarre thought experiments play in our reasoning more generally.² It is not, as some would have it, that there is a level or degree of outlandishness beyond which we should venture, but rather that the admissibility of the outlandish depends on the argumentative context—or so I will argue.

2. The standard analysis of non-identity and its relevance to applied ethics

The non-identity problem consists in the claim that determining whether or not a harm has been done to a particular person must be constrained by considerations of whether less harmful alternative courses of action would have led to that person existing. Parfit’s description of the problem in *Reasons and Persons* begins with the claim

¹ Jakob Elster explores the admissibility of the outlandish in his 2011 article. See also Wilkes 1988 and Pogge 1990.

² Herein I argue for thought experiments having a variety of roles in philosophical reasoning. However, writers in this area typically pick out only one such role; so, for instance, Elster (2011) treats them as primarily means of generating intuitions for testing our moral principles.

that whether a particular person exists is dependent (at least in part) on when exactly the procreation takes place.

The 14-Year-Old Girl: This girl chooses to have a child. Because she is so young, she gives this child a bad start in life. Though this will have bad effects throughout the child's life, his life will, predictably, be worth living. If this girl had waited for several years, she would have had a different child, to whom she would have given a better start in life. (Parfit 1984: 358).

However, Parfit asserts that in assessing the relative harm of the circumstances of any birth, one cannot *compare them with other scenarios* involving alternative policies and actions in which one would not have been born at all. Thus the child cannot be said to be harmed by his mother's action of conceiving him at fourteen, for the very reason that he would not have been born at all if she had conceived later in life.

Notice that what we are concerned with here is one type of *counterfactual or modal harm*. It is not what the girl did to her child directly that is under scrutiny but what she didn't (namely not having him when she was eighteen). The focus here is on what an agent could have done which was less harmful than what he or she did in fact do. We might call this "could-have-done-better" counterfactual harm. We can contrast this with cases where we might hold someone responsible for some event that they were lucky did not occur, even though it might well have. For instance, if I run through a series of red lights without stopping, my action would be held to be morally blameworthy even when no harm comes to me or anyone else. This we might call 'there-but-for-the-grace-of-god' harm. In this paper it is the former type of harm and consequent blame with which we are concerned.³

Parfit raises the non-identity problem to argue for what he calls *impersonal* as opposed to *person-affecting* moral frameworks. He claims that we should reject the view that an outcome can only be worse if it is worse *for someone* and, further, that we act wrongly only by making a *particular* existing person worse off. Instead we have an obligation to do what would maximise overall happiness rather than the happiness of particular people.

In recent years the problem has taken centre stage in debates in applied ethics concerning the use of new reproductive technologies, whether in some cases one would be better off not being born and environmental debates about our obligations to future generations.⁴ In the area of bioethics, it is commonly invoked in discussions concerning the harms, benefits and duties associated with the manipulation and alteration of the genetic make-up of future individuals. If we transform the genetic make-up of a future being in such a way that it is no longer the same individual, do we harm the child who would have been born had

³ Whilst this might seem a little odd, it is quite common to define harm in terms of counterfactuals. See for instance, Feinberg 1986.

⁴ See, for instance (as just a glimpse into this vast literature) Roberts and Wasserman 2009 and Archard and Benatar 2009.

we not interfered? And who do we benefit by such interference? There is also a debate in the bioethical literature about wrongful life in which the non-identity problem has had a significant role. (Archard 2004) How can an individual be said to be harmed by being brought into existence when the contrast is with a state of affairs in which they do not exist and hence one cannot compare their relative state of well being?

In the area of environmental ethics, writers explore the implications of choosing conservation or degradation of the environment (Carter 2001). Suppose we have a choice between these two outcomes. Whichever choice we make has consequences, not only for the quality of life of those who inhabit the future, but the very identity of those future citizens. Suppose we choose degradation. Clearly the lives of those who inhabit a degraded environment are worse than those who inhabit a non-degraded one. But if they would not have existed in the non-degraded environment, and their lives are still sufficiently good that they are better off being born than not being born at all, then can we say that they are some how harmed by our choosing the degraded option?

3. *Non-identity and thought experiments*

The bulk of the specifically philosophical discussion of the non-identity problem has focused on whether one should indeed become an ‘impersonalist’ or whether there might be a rights-based solution to the problem. However, these competing responses to the problem all appear to accept implicitly a thesis upon which I wish to place some pressure: namely that nomic claims about identity should constrain the kinds of imaginary scenarios which we might employ to judge the relative welfare of a person.⁵

To see this let us reconsider our initial scenario regarding the fourteen-year-old mother and compare it with the following case.

Kangaroo-like Gestation Imagine that human beings had similar reproductive capacities to those of kangaroos. Kangaroos have the ability to delay gestation of fertilised embryos, during lean periods, until such time as conditions are suitable for the bearing of healthy offspring. If human beings were like kangaroos then it would be possible for the fourteen-year-old mother to delay the birth of her child until she was eighteen, thus ensuring a better life for her child.

In this case it would be the identical child that is born four years later.

⁵ In the entry on the Non-identity problem in the *Stanford Encyclopedia of Philosophy*, M.A. Roberts notes in passing that in these cases the range of courses of action must be consistent with existing medical and genetic technologies. Later Roberts notes that since Parfit first described the case reproductive technologies have advanced in a way that makes it, at least, theoretically possible for the 14 year-old girl to the same child as the later born child. Nonetheless Roberts claims that probabilities need to be taken into account here and, even with new technology, it is *very probable* that the two children would be non-identical.

Accordingly, in this scenario if she chose to have the child at 14 when she could have had the identical child at say 18, we would much more willing to say that she has harmed that particular child.

Far less fancifully, we can readily imagine technology that would allow us to take fertilised embryos from the wombs of recently impregnated women and freeze them in a storage facility where they would be kept until such time as the woman was ready to gestate the future child. So here again our fourteen-year old girl would be able to bear the identical child in circumstances more conducive to his over-all wellbeing.

There are also cases we could consider that are much closer to the kinds of worries about identity we find in the bioethics literature, where the harm involves some genetic disadvantage. To see this consider another scenario:

The Bridesmaid's Dress: a woman knows that if she conceives in the present month, the child conceived will suffer some serious genetically-based ailments, but not so severe that the child would be better off not existing. However, if she delays conception she will not fit into her bridesmaid's dress at wedding scheduled for nine and half months time. The woman decides to conceive and gives birth to a child.

Following the reasoning of the non-identity problem we cannot say that the child is harmed by comparing his welfare with that of a hypothetical child born a month later, since our knowledge of biology leads us to the belief that it would not be the same child. The standard and quite sensible view is that different eggs and different sperm make for a different genetic identity. Moreover, genetic identity is assumed to be a necessary, if not a sufficient, condition of personal identity: that is to say, one cannot be the same person without being genetically identical.⁶

But it is, of course, possible to imagine cases whereby the child who is born a month later would in fact be genetically identical. The woman need only be capable of delaying gestation for the later child to be genetically identical. What these kinds of cases demonstrate is that the non-identity of, for instance, later born children is not a *necessary* truth about the world, for there are possible cases where the child born a month later or four years later, or even a hundred years later would be the same person or at least genetically identical. The child born of the woman at the age of fourteen and the child born four years later are not *necessarily non-identical*.

In order to rule out such counter-examples, advocates of the non-

⁶ In saying this we need to acknowledge explicitly the further problem concerning the role of genetic identity in personal identity: merely being genetically identical does not make an agent the identical person. However, for our purposes we can put that question to one side since the thought in the non-identity problem is simply that we do not have genetic identity in later born children and hence no claim of harm can be made in relation to such children.

identity problem must deem inadmissible, when determining relative welfare, any counterfactual states of welfare found in conceivable worlds that defy our current views of what is *possible in the circumstances* and which will usually be covered by the nomically possible. It is not simply *non-identity* that is determining whether or not harm has occurred in our world, but non-identity in circumstances closely resembling ours. Let us call this the nomic constraint on the use of thought experiments in the assessment of harm (NC). We might formulate it as follows:

The Nomic Constraint: In assessing the relative moral status of any action we should not compare it with counterfactuals that are nomically impossible or are impossible in the present circumstances.

Another way of putting this point is that the non-identity problem rules out certain ‘modal harms’ in the assessment of relative welfare. When we decide whether a person is harmed by our choosing (or not choosing) a particular course of action, their well-being in possible worlds in which, given our current levels of understanding and technology or given our understanding of the laws of nature, they could not exist, are not relevant.

This is not uncontroversial. Philosophers, especially in the area of metaphysics, make regular use of examples that defy the laws of physics, as we know them, whilst others—who are in a minority—reject their employment. It is useful here to distinguish, roughly following Elster (2011: 242), between *conceivabilists* and *realists*.⁷ According to the conceivabilist, so long as a case is conceivable then it is legitimate to use it as a means of testing our theories. According to the realist, on the other hand, only cases that could plausibly occur given the world as it is should be used to test our theory. In the area of moral philosophy this view is sometimes defended on the grounds that since moral principles are meant for guiding action in the world, cases drawn from other worlds are irrelevant. Kathleen Wilkes (1988) makes a similar point in relationship to questions of personal identity. Other realists argue that we lack the cognitive capacity to apply our intuitive faculties to outlandish cases (Elster 2011: 242).

Clearly the nomic constraint fits broadly within realism. Curiously, given the context of our discussion, Parfit’s general view on the matter would appear to be essentially conceivabilist. In *Reasons and Persons*, in the midst of a discussion of whether Nozick’s imagined Utility Monster is deeply impossible (and hence irrelevant to moral debates on public policy), Parfit notes that ‘even an impossibility may provide a test for our moral principles’. He claims that “[W]e cannot simply ignore imagined cases.” (Parfit 1984: 389). Yet, the non-identity problem

⁷ In the main Elster treats imaginary cases as being used to elicit intuitions to test our moral principles, a claim which I will challenge later in the paper. However, at various points he does acknowledge different roles that thought experiments might play.

only generates the problems it does if we rule out the kinds of imaginary cases outlined above.

Whatever one's stance in the debate between realism and conceivability, it is somewhat anomalous given the regular use of modally outlandish cases by philosophers, that those exploring the non-identity problem have accepted so readily the idea that genetic identity determines personal identity and hence constrains what might count as a possible form of harm.

Let us summarise the line of reasoning thus far. According to Parfit non-identity means that the child born of the 14 year-old girl cannot complain of some harm being done to him or her. It is not the case that Parfit claims there is no harm; instead it is harm in terms of the overall sum of welfare, rather than harm to any particular single individual. However, the non-identity that underpins this argumentative move is—as the case of the kangaroos demonstrates—*contingent not necessary*. In focusing solely on *contingent non-identity* we must rule as inadmissible evidence any counter-examples from far-off possible worlds. This mode of ruling inadmissible I shall refer to as the 'nomic constraint'.

4. *The Non-identity problem and modal moral luck*

It would appear then that those debating the implications of the non-identity theory are agreed on one point at least, namely that nomic claims about identity should constrain the kinds of imaginary scenarios that we can legitimately employ to judge the relative welfare of a person. Let us focus a little more closely on this feature of the debate. What we have here is a *nomic constraint* on the imaginary scenarios and thought experiments that will count as relevant cases in the assessment of individual harm. The suggestion is that we cannot employ nomically impossible examples to determine the moral status of action or future policy and hence our use of thought experiments in moral philosophy will be constrained here by the extent to which those thought experiments introduce cases which are nomically possible.

One further interesting consequence of this is that it appears to entail a commitment to the idea of what we might call "modal moral luck", according to which the moral status of an action is contingent upon which particular possible world one inhabits. The term is related to the idea of 'moral luck' that Bernard Williams famously proposed to cover cases where the moral status of an action or even a whole life depends upon how things turn out. Moral luck describes cases where "an agent can be *correctly* treated as an object of moral judgement", despite the fact that a significant aspect of what that agent is assessed for "depends on factors beyond his control" (Nelkin 2004). In the Kantian tradition matters are entirely different for there it is assumed that we are only judged in terms of the motives with which we act. Our

intuitions might well be thought to support the Kantian view since the very idea of moral luck sounds oxymoronic. Further it seems unfair to be judged by circumstances that are outside of our control. Yet at the same time our ordinary moral thinking seems to suggest that there is such a moral phenomenon. As Thomas Nagel notes:

Where a significant aspect of what someone does depends on factors beyond his control, yet we continue to treat him in that respect as an object of moral judgement, it can be called moral luck. (Nagel 1979: 59)

In order to illustrate his point, Williams tells the story of the painter Paul Gauguin. Gauguin left his wife and family in France to pursue his career as a painter. Williams' point is that if it had turned out that Gauguin was a mediocre painter rather than a gifted artist, then we would judge his life rather differently. The general idea then is that we are more hostages to moral fortune than the standard Kantian analysis would have us believe.

Modal moral luck, by way of contrast, involves cases where the moral assessment will differ according to what world, of all those possible, it is in which the action occurs. Thomas Nagel raises a related (although far less extensive) idea when he discusses the category of 'circumstantial luck' (Nagel 1979). Circumstantial luck is luck about the circumstances in which one finds oneself. Nagel's example concerns Nazi collaborators in Germany during the period of the Third Reich. We condemn them for the morally appalling acts they performed, but if they had been shifted to South America in, say, 1929, perhaps they might have led morally exemplary lives. If we provide a different moral evaluation of the lives of the Nazi collaborators with their hypothetical counterparts in South America, then we have a case of what Nagel calls circumstantial moral luck. Modal moral luck extends the idea of relevant hypothetical counterparts to a much wider range of possible worlds, including those that we might think of as being nomically impossible.

According to the idea of modal luck then the moral status of one's action is contingent upon what possible world we inhabit. To illustrate the idea let us return to Parfit's original example of the fourteen-year old girl. Does the fourteen-year old girl harm her son by having him when she is fourteen? Is he harmed by not being born four years later? On this line of reasoning the moral status of her action depends upon what possible world it is that she inhabits. In the world where human beings cannot delay gestation of fertilised embryos then she does not harm him because he would not exist. In the world where human beings can perform this procreative trick, then she does harm him. The moral status of her action is thus contingent upon which possible world she inhabits. It is, we might say, a matter of modal moral luck whether she can be said to have harmed her son.

Modal luck would cover what we might call 'epistemic cases' where our lack of knowledge makes an outcome impossibly remote for us. There will be cases where our ignorance makes it impossible for us to

take advantage of some harm-lessening process. Imagine that if we drink the right admixture of iodine and calcium that we can delay gestation for up to ten years. In this case again the fourteen year old could take this concoction and give birth to the same child four years later when she is better placed to raise him. In a world where this technique is widely known we might well judge the action of the woman of not delaying the pregnancy quite differently from we would in a world in which this information was not available. Again the fact that the same action is evaluated quite differently is a matter of modal moral luck.

Modal luck has interesting implications, not merely for debates regarding how we might determine whether harm has occurred and what kinds of counterfactual considerations are relevant, but also for our use of thought experiments in ethics more generally. Think about this in relationship to one of the more notorious thought experiments in the ethical literature, Michael Tooley's case of the superkittens (Tooley 1983). Tooley's target was the potentiality principle that is often used to oppose abortion. In attacking the idea that it is wrong to kill fetuses because they are potential persons, Tooley employs a thought-experiment involving highly rational cats to generate a putative *reductio ad absurdum*. He writes:

Suppose that at some time in the future a chemical is discovered that, when injected into the brain of a kitten, causes it to develop into a cat possessing a brain of the sort possessed by normal adult human beings. Such cats will be able to think, to use language, to make decisions, to envisage a future for themselves, and so on—since they will have all of the psychological capacities possessed by adult humans. If one maintains that it is seriously wrong to kill adult members of the species *Homo sapiens*, one must also....hold that it would be seriously wrong to kill cats that have undergone such a process of development. (Tooley 1983: 191)

He then claims that it follows that it is *prima facie* no more seriously wrong to kill a human organism that is a potential person, but not a person, than it is intentionally to refrain from injecting a kitten with the special chemical, and to kill it instead. This he suggests shows why the potentiality principle is wrong for in the example provided above it would be intrinsically wrong to refrain from injecting a kitten and killing it instead. The claim that it is *intrinsically* wrong to kill a foetus is therefore putatively reduced to absurdity.

However, we should be extremely wary of accepting this thought-experiment as a *refuter*. To do so would be to overlook the fact that we are making decisions in this world and the contingent facts about how this world is actually constructed are relevant to how we frame our decisions. In the possible world in which we possessed such a chemical, refraining from injecting a kitten and subsequently killing it *would be morally equivalent* to killing a foetus (and that says nothing about whether or not it would be intrinsically wrong). We might think of this as a form of *modal moral luck*. For the kitten-killer it will be a matter of 'modal moral luck' whether or not he inhabits a possible world

in which killing kittens is equivalent to killing foetuses. But in our world they are not morally equivalent along the lines of comparison that Tooley discusses.

We can restate the point in the following way. Imagine two different people in two different worlds, World A and World B, who both accidentally run-over a young kitten. In world A we have the drug to turn cats into supercats whereas in B we do not. In the world of potential supercats the action has more bad-making features than in world B. The actions may well be equivalent in terms of their responsibility, causality and intention. But we might say that killing A is morally worse and, that it is so for the driver concerned, is a matter of modal moral luck.

5. *Modal Constraints and the Assessment of Harm*

However, nomic constraints on counterfactuals are decidedly odd in philosophy. How far would we want to generalise this ‘nomism’? Would we want to adhere to it as a general principle, for bizarre counterfactuals are commonplace in philosophical debate. Think, for instance, in metaphysics of discussions of the swampman who emerges out of the swamp with all of the same properties as an ordinary human.⁸ Equally, in the philosophy of mind there has been a great deal of discussion of zombies who behave as if they are conscious but in fact are not.⁹

If we turn our attention to ethics and applied philosophy again we find considerable use of what one might think of as ‘bizarre examples’. Think, for example, of Judith Jarvis Thomson’s case of the people-seeds that she raises when discussing the morality of abortion (Thomson 1971). These people-seeds float around in the atmosphere and if one does not place the appropriate guards on one’s windows then they will float inside one’s house, attach to the carpet and begin growing into people.

The use of bizarre examples, then, is standard-practice in many areas of philosophy in general and many would regard it as intellectually productive. Although there are realist critics of the modally bizarre—most notably Kathleen Wilkes and Thomas Pogge—the demand for constraints in terms of nomic possibility is somewhat anomalous. Realism of this variety would be thought by many as stymieing philosophical analysis and thus would seem to require justification.¹⁰ The default position within philosophy is surely that the modally bizarre are admissible.

A further possible problem with *the nomic constraint* concerns its implications for commonly employed notions in moral philosophy such as *universalizability*. Does it restrict the circumstances in which it is

⁸ See, for instance, Davidson 1987 and Millikan 1996.

⁹ For useful surveys of this literature see Kirk 2005, Marcus 2004, Cottrell 1999 and Dennett 1995, 1991.

¹⁰ See Wilkes 1988 and Pogge 1990.

possible for me to imagine myself? Universalizability involves (roughly speaking) the idea that in determining the moral status of a proposed action we need to ask ourselves firstly what the world would be like if everyone did it and secondly how would we feel if it were to be performed on us. Any action to which we would not assent in these hypothetical circumstances is morally impermissible. The nomic constraint is unlikely to create difficulties for the first element since universalising some action will rarely require invoking impossible states of affairs. However, problems do arise when we consider the second condition regarding thinking oneself into the shoes of another. If I am supposed to imagine what it would be like to be a member of a very different culture and imagine how I would feel about certain forms of discrimination, then this seems to require that I imagine something impossible, namely that I am a very different person than I am. If non-identity rules out children born four years later, then *a fortiori* it might seem to rule out such radical transformations. But perhaps this is a red herring since the non-identity problem concerns what are possible counterfactual conditions for a particular person, not how I might imagine myself. Be that as it may, the significant point for our purposes is that, in determining relative welfare, the non-identity problem deems inadmissible any harm found in conceivable worlds that defy our current views of the nomically possible. However, if this were to be taken as a more general principle in philosophy, would it 'poison the well' of moral theory, since it would undermine our use of thought experiments?

For realists, however, who wish to eliminate the use of the outlandish these are not costs but desirable outcomes. Yet, the view is controversial. Must anyone who regards the non-identity as raising genuine philosophical issues about the nature of harm thereby commit themselves to realism about thought experiments?

Fortunately, this would appear not to be the case. It need not have such extreme ramifications for philosophical method if we limit the scope of the nomic constraint that underpins it. Reconsider the topic out of which the non-identity problem arose. It concerns harm and, more specifically, *actionable* harm. In this case we are concerned with assigning responsibility—and perhaps blame—for harms done to future people by dint of the timing of their birth or their genetic make-up. In pursuing such questions we are constrained by what could *reasonably be expected* in the current circumstances, not by what is logically or metaphysically possible. Harm, in this context, is a practical notion that does not need to be evaluated in relation to every possible world.

The upshot is that we need to revise our view about what assumptions underpin the non-identity problem. Our original nomic constraint advised that when assessing the relative moral status of any action, we should not compare it with counterfactuals that are nomically impossible or are impossible in the present circumstances. However, in order to generate the non-identity problem this constraint need not be generalised. Relying on the nomic constraint in the context of debates

about harm need not commit one to it as a more general principle for moral theory, or indeed for philosophical theory more generally. Cast in a more modest way, then, the constraint can be rewritten as follows:

A modal constraint with restricted scope: In assessing the relative moral status of any action *in terms of actionable harm* we should not compare it with counterfactuals that are nomically impossible or impossible in the present circumstances.

This weaker version of the constraint on imaginary cases, unlike stronger forms of realism, does not have such serious implications for philosophical method since it allows for the continuation of bizarre thought experiments in general, but rules out specific cases where there is actionable harm. It rules out ‘kangaroo-style’ cases when discussing harm in the current context but does not rule out counterfactual forms of harm, such as in the case of the inebriated driver who fortunately harms nobody, from being relevant when assessing actionable harm. Crudely speaking, drunk drivers are in and kangaroos are out.

There remain, we must admit, vexed questions of how to determine what might be impossible in the current circumstances. There are two closely related parts to this difficulty. The first concerns how we operationalize the constraint. If one is assessing harm—and this is particularly true if that harm were to have any legal ramifications—how can one be sure what really is impossible? There is always the possibility that some states of affairs are in fact possible despite our ignorance of that fact. Secondly, rapid technological changes can mean that what had been difficult even to imagine at one point in history can change overnight. Nonetheless, as troubling as these considerations of the assessment of harm might be, they do not undermine the general point about the legitimacy of ruling some thought experiments inadmissible on such grounds.

6. *Modal constraints and the argumentative context*

Thus far I have argued that the modal constraints associated with the non-identity problem are justifiable because we are concerned with actionable harm. At this point I wish to draw a more general lesson about the admissibility of outlandish thought experiments. Some critics, such as Elster, focus primarily, when assessing admissibility, on the modality of particular thought experiments. However, I beg to differ. Outlandishness or bizarre modality is not the central issue in relation to the admissibility of thought experiments, rather it is the argumentative context. It is not the violation of what is *possible alone* that justifies rejecting the claim that the child is harmed by not being born four years later. Instead it is the violation of what is possible *in an argumentative context in which such violations matter* that will justify constraint. ‘Modal violation’ is but one possible relevant factor when assessing such admissibility but there will be many others. The general idea is that constraints can be justified in relation to the argumentative context.

The approach I sketch briefly below involves what we might call ‘thought experimental pragmatics’, in which the admissibility of a thought experiment will be determined not by the specific features or inherent modal nature of the thought experiment itself, but by the role that it plays in the argumentative context. We can illustrate this by drawing an analogy with pragmatics in linguistics and the philosophy of language (Lycan 1995). There the term ‘pragmatics’ refers to approaches that explore the contextual dimensions of our use of language and the context-dependence of various features of linguistic interpretation (Korta and Perry 2011). The analogous idea proposed here is that the need for modal constraints should be determined with reference to the argumentative context.

In order to explicate this approach I begin by considering the diverse roles that thought experiments play in argumentation. There are, I suggest, a number of distinct functions that thought experiments play in arguments, which importantly cannot be reduced to a single function. Below I identify three such functions; although the list is not intended to be exhaustive. One consequence of this approach is that there is unlikely to be a single characterisation of what makes for a successful thought experimentation. (Here I have in mind the kind of regimentation that Soren Haggqvist gives in his 1996 book *Thought Experiments in Philosophy*). Note also the difference with Roy Sorensen’s account in *Thought Experiments* (1992) in which Sorensen argues that the single role of thought experiments is to test modal consequences. Sorensen leaves this conception of function with the comment that the ‘apparent narrowness of its function eases once we realize that there are many kinds of necessity’ (Sorensen 1992: 6). Jacob Elster notes that imaginary cases are used in ethics to elicit intuitions against which moral principles might be tested and presumably elsewhere in philosophy they are used to challenge other principles (Elster 2011: 241), be they metaphysical or epistemological and so on.¹¹ Again he only mentions one function. However, on the account suggested herein, there is no such single function, no matter how broadly construed.

There are, then, at least three roles that thought experiments play in philosophical arguments. First, some thought experiments function as counter-examples in philosophical disputations. In responding to a theory or a definition that is intended to be either necessarily or universally true one might attack the position either by providing a counter-example or by demonstrating that the theory has absurd consequences.¹² The use of such refuters is a commonplace in moral philosophy. For example, in responding to the claim that it is *always* wrong to lie, an opponent of this view might argue that at least some lies are per-

¹¹ As the article proceeds, Elster does mention various uses of imaginary cases that do not fit quite so neatly into this understanding of their function.

¹² Roy Sorensen refers to these thought experiments as ‘refuters’, although, as he notes, not all refuters involve thought experiments (see Sorensen 1992: 153).

missible and do so by providing a case where most people would admit lying was acceptable (such as when telling a so-called ‘white lie’). The second category of thought experiments involves what we might refer to as ‘intuition pumps’. This is a term that is used in a variety of ways by philosophers—sometimes as a synonym for thought experiments and sometimes to refer to what the author believes is a pernicious mode of reasoning—but one common usage is where it refers to a class of thought experiments that aim to lead us, via our reactions to a single thought experiment, towards some general kind of conclusion. Trolley problems might be a case in point. Here we are meant to infer from the fact that we would choose to save the five rather than the one person on the track that numbers do count morally.¹³ A third category of use to which thought experiments are often put by philosophers is as ‘clarificatory devices’ where the aim is to clarify our views on a difficult topic. Perhaps the most widespread of these are the ‘commitment cleavers’, that is, cases where thought experiments are used to enhance our understanding by teasing apart distinct, but easily conflated, principles. Presumably this is at least part of the significance of the story in the *Republic* of the Ring of Gyges that Plato raise in the midst of a debate about the nature of justice (Plato 1974: 36). The tale of this ring enables us to distinguish between those who endorse justice on the grounds of prudence and those who do so because they regard acting fairly to be a fundamental moral obligation that holds regardless of any benefits that might or might not accrue from being seen to act justly. Through Plato’s use of this dramatic device, the interlocutors in the *Republic* are forced to be more specific about their ethical and political commitments. In the end is it mere prudence or our intrinsic duty to act justly that underpins their publicly avowed commitment to justice?¹⁴ Sometimes these argumentative devices are designed so as to test or clarify what a theory—as opposed to a person—might be committed. One might, for instance, devise a thought experiment that illuminates the difference between a Kantian and a Utilitarian approach to moral issues.

With this taxonomy in mind—and, more importantly, being apprised of the thesis that thought experiments play a variety of roles in argumentation—let us now return to the issue of method in ethics and how the idea of modal constraints relating to possibility might best be understood. (My assumption continues to be that Parfit’s problem raises important questions for ethical methodology). The claim herein is that the thought experiments philosophers often use in ethics can—

¹³ One difficulty with attempting to use thought experiments in this manner is that people’s responses to thought experiments are often surprisingly varied. Hence they do not always pump our intuitions in the direction that the interlocutor hopes or expects.

¹⁴ As C. L. Ten notes such thought experiments help us to determine whether a particular principle or commitment is “fundamental or subordinate” (Ten 1987: 21).

and should—be constrained, if the argumentative context renders more modally extravagant cases to be irrelevant. To illustrate how this would work, consider the following cases. If, for example, we were debating the rightness or wrongness of incarcerating human beings without trial it would be inadmissible in this context to introduce, as a relevant case, the example of possible people who enjoy being incarcerated. Given the practical context, such beings would not provide counter-examples to claims that it is wrong to imprison without trial. What matters here is not so much the *content* of the example in and of itself, but rather the *context* of the topic under discussion. With respect to the non-identity problem the argumentative context involves actionable harm). Yet, there will be other cases in which we are considering general principles of morality where it might well be appropriate to raise the moral consequences of persons who enjoy being incarcerated.

Consider a further example. Suppose that two philosophers are debating the issue of whether or not a pregnant women should have more say in any decision about whether or not to continue with a pregnancy than the man who fathered the foetus. Now suppose that in the context of this debate one of the disputants (let's call him 'Jim' for the sake of the example) raises the following imaginary scenario.

Ectogenetic birth: Imagine a world in which human beings can be gestated entirely outside of the womb, in perhaps an incubator of some kind. In such a world there is no reason to think that the female parent has more rights than the male parent in determining whether the gestation should continue.

Jim argues that this example undermines the claim that women should have greater say than men in the determination of whether gestation should continue. The idea would be that the moral claim of women having greater rights here does not hold in all possible worlds for the very reason that it is not true all possible worlds that women will be responsible for the gestation. So far so good, but in line with the preceding dictum about context (i.e. our modal constraint), our friend Jim cannot use the case to make claims about whether or not women should have more say *now* in circumstances where women are in fact responsible for gestation. In this case, contingent features of the problem—namely the fact that we do not currently have ectogenetic birth—are relevant to our moral deliberations and should not be over-ruled by merely logically possible cases. And again it is a matter of modal moral luck whether or not a man might be thought to have less rights in determining whether gestation should continue.

The more general lesson here is not that we should always constrain our thought experimentation by contingent empirical realities, nor that bizarre examples should be expunged from moral thinking but, simply, that there will be cases—particularly in applied ethics and political philosophy—where the argumentative context is such that assessment of contingent factors about *what is actually possible* matters.

It is the argumentative context not the outlandishness of any thought experiment that should be our primary concern in determining admissibility. We must not rule out outlandish experiments merely because they are outlandish.

7. *Concluding remarks*

The non-identity problem raises difficult questions about whether it is possible to harm another person who, if we had not acted as we did, would not have existed. A significant assumption here is that if the harmful action had not been undertaken then the person would not exist. Yet, in each of the cases raised, it is entirely possible to imagine future scenarios—which are admittedly outlandish—in which the person existed without the harm. In order to sustain the claim that we cannot harm such individuals, one must rule out such imaginary scenarios. But why rule out these imaginary cases? My aim was to discover what might justify such constraints that are curious given the kinds of scenarios regularly explored in philosophical debate.

After considering the ‘*nomic constraint*’ (which would rule out all modally bizarre scenarios) I proposed, by way of justification, a more moderate constraint that focuses on the fact that in this instance we are concerned with actionable harm and hence the kind of scenarios that are relevant should be restricted by that concern. Scenarios are restricted in the non-identity problem because of the argumentative context. This is the core idea of the approach of ‘*thought experiment pragmatics*’.

The more general claim I made is that it is the argumentative context, not the modal content of a particular imaginary case, which determines whether it is admissible. The approach defended here is midway between, on the one hand, those who would adopt an ‘*anything-goes*’ policy and, on the other, those like Kathleen Wilkes who regard outlandish thought experiments as intellectually pernicious (Wilkes 1988). There are, I would suggest, genuine grounds for limiting in certain debates the range of thought experiments that are to be regarded as admissible. This is an important lesson for areas of applied philosophy. While matters are somewhat different when considering, for instance, general moral principles, when the issue is very much a practical question of applied philosophy, such as is the case with the assessment of responsibility and actionable harm, it is appropriate to limit the range of relevant cases. The default position should be that outlandish examples are admissible until such time as an interlocutor can demonstrate that, given the context, they are irrelevant to the debate at hand.

References

- Archard, D. 2004. "Wrongful Life." *Philosophy* 79: 403–420.
- Archard, D. and Benatar, D. (eds.). 2009. *Procreation and Parenthood: The Ethics of Bearing and Rearing Children*. Oxford: Oxford University Press.
- Brock, D. 1995. "The Non-Identity Problem and Genetic Harms: the case of Wrongful Handicaps." *Bioethics* 9: 269–275.
- Carter, A. 2001. "Can We Harm Future People?" *Environmental Values* 10: 429–454.
- Cottrell, A. 1999. "Sniffing the Camembert: on the Conceivability of Zombies." *Journal of Consciousness Studies* 6: 4–12.
- Davidson, D. 1987. "Knowing One's Own Mind." *Proceedings and Addresses of the American Philosophical Society* 60: 441–458.
- Dennett, D. 1995. "The Unimagined Preposterousness of Zombies." *Journal of Consciousness Studies* 2: 322–26.
- Elster, J. 2011. "How Outlandish Can Imaginary Cases Be?" *Journal of Applied Philosophy* 28 (3): 241–258.
- Feinberg, J. 1986. "Wrongful Life and the Counterfactual Element in harming." *Social Philosophy and Policy* 4: 145–177.
- Gendler, T. S.. 2000. *Thought Experiments: On the Powers and Limits of Imaginary Cases*. New York: Garland Publishing.
- Hare, C. 2007. "Voices from Another World: Must we respect the Interests of People who Do Not, or will Never Exist." *Ethics* 117: 498–523.
- Haggqvist, S. 1996. *Thought experiments in philosophy*. Stockholm: Almqvist and Wiksel International.
- Harris, J. 2000. "The Welfare of the Child." *Health Care Analysis* 8 (1): 27–34.
- Holtung, N. 2001. "On the Value of Coming into Existence." *The Journal of Ethics* 5: 361–384.
- Kirk, R. 2005. *Zombies and Consciousness*. Oxford: Clarendon Press.
- Lycan, W. 1995. "Philosophy of Language." In R. Audi (ed.). *The Cambridge Dictionary of Philosophy*. Cambridge: Cambridge University Press: 586–589.
- Korta, K. and Perry, J. 2011. *Critical Pragmatics: An Inquiry into Reference and Communication*. Cambridge: Cambridge University Press.
- Marcus, E. 2004. "Why Zombies are Inconceivable." *Australasian Journal of Philosophy* 82: 477–490.
- Millikan, R. 1996. "On Swampkinds." *Mind and Language* 11 (1): 70–130.
- Nagel, T. 1979. *Mortal Questions*. New York: Cambridge University Press.
- Nelkin, D. K. 2013. "Moral Luck." In *The Stanford Encyclopedia of Moral Philosophy* (retrieved August 2013) <http://stanford.library.usyd.edu.au/entries/moral-luck>.
- Parfit, D. 1984. *Reasons and Persons*. Oxford: Oxford University Press.
- Plato. 1974. *The Republic*. Translated by D. Lee. Harmondsworth: Penguin.
- Pogge, T. 1990. "The effects of prevalent moral conceptions." *Social Research* 57 (3): 649–663.
- Sorenson, R. 1992. *Thought Experiments*. New York: Oxford University Press.

- Roberts, M. A. 1998. *Child versus Childmaker: Future Persons and Present Duties in Ethics and the Law*. Lanham: Rowman and Littlefield.
- Roberts, M. A. 2013. "The Nonidentity Problem." In *The Stanford Encyclopedia of Moral Philosophy* (retrieved October 2013). <http://stanford.library.usyd.edu.entries/nonidentity-problem>
- Roberts, M., and Wasserman, D. (eds.). 2009. *Harming Future Persons: Ethics, Genetics and the Nonidentity Problem*. Dordrecht: Springer.
- Ten, C.L. 1987. *Crime Guilt and Punishment*. Oxford: Clarendon Press.
- Thomson, J. J. 1971. "A Defense of Abortion." *Philosophy and Public Affairs* 1 (1): 47–66.
- Tooley, M. 1983. *Abortion and Infanticide*. Oxford: Clarendon Press.
- Wilkes, K. 1988. *Real People: Personal Identity without Thought Experiments*. Clarendon Press: Oxford.
- Woodward, J. 1986. "The Non-identity Problem." *Ethics* 96: 804–831.
- Wrigley, A. 2006. "Genetic Selection and Modal Harms." *The Monist* 89 (4): 505–525.