

Staking our future: deontic long-termism and the non-identity problem

Andreas Mogensen

Global Priorities Institute | September 2019

GPI Working Paper No. 9-2019



Staking Our Future: Deontic Longtermism and the Non-Identity Problem

Andreas L. Mogensen

Global Priorities Institute, Faculty of Philosophy, Oxford University

ABSTRACT: Greaves and MacAskill argue for *axiological longtermism*, according to which, in a wide class of decision contexts, the option that is *ex ante* best is the option that corresponds to the best lottery over histories from *t* onwards, where *t* is some date far in the future. They suggest that a *stakes-sensitivity argument* may be used to derive *deontic longtermism* from axiological longtermism, where deontic longtermism holds that in a wide class of decision contexts, the option one ought to choose is the option that corresponds to the best lottery over histories from *t* onwards, where *t* is some date far in the future. This argument appeals to the *Stakes Principle*: when the axiological stakes are high, non-consequentialist constraints and prerogatives tend to be insignificant in comparison, so that what one ought to do is simply whichever option is best. I argue that there are strong grounds on which to reject the *Stakes Principle*. Furthermore, by reflecting on the Non-Identity Problem, I argue that there are plausible grounds for denying the existence of a sound argument from axiological longtermism to deontic longtermism insofar as we are concerned with ways of improving the value of the future of the kind that are focal in Greaves and MacAskill's presentation.

I.

How should we choose between the present and the future? Some of the most pressing problems that we face are fleeting when viewed in historical perspective. Millions currently suffer in extreme poverty. However, the percentage of people living in extreme poverty fell dramatically near the end of the 20th century, virtually halving between 1990 and 2010. It could well fall to zero within our lifetimes (The Economist 2013). Other problems that press upon us will spread across

the very long run. Climate change is the most salient example. CO₂ released into the atmosphere today has the potential to affect the climate for thousands of years into the future (Archer et al. 2009). In deciding which problems to prioritize, should we be more concerned about the present or the future?

Hilary Greaves and William MacAskill (2019) argue that the future should typically have priority. More specifically, they argue for a thesis they call *axiological longtermism*. Axiological longtermism is a hypothesis about so-called *ex ante* axiology, which involves ranking probabilistic prospects as better or worse (Broome 1991). Any such prospect may be called a *lottery*. In their formulation, axiological longtermism says that in a wide class of decision contexts, the option that is *ex ante* best is the option that corresponds to the best lottery over histories from *t* onwards, where *t* is some date far in the future. Thus, we could pick *t* to be the year 2312. Axiological longtermism tells us that if we are comparing different options and we want to rank their associated prospects as better or worse from the moral point of view, we can in many cases effectively ignore possible outcomes associated with the years 2019-2312 and rest our comparison entirely on the probability distribution over possible outcomes associated with the oceans of time that fall between the year 2312 and the heat death of the universe.

The argument that underpins this view is easy to grasp. When we evaluate acts in terms of their consequences, all of their consequences should be counted, no matter how distant in space or time. Furthermore, the long-term future is potentially vast. There is so much more time between 2312 and the heat death of the universe than there is between 2019 and 2312. There are so many more people who could populate that time. It should therefore be easy to see why we might think that the potential effects of our actions distributed across those gigantic possible future populations should carry more weight than their near-term effects, which concern far fewer people.

The action that is associated with the best prospect need not be considered obligatory (Kamm 2000). We may believe, for example, that the ends do not always justify the means: that certain outcomes, although they would be better as outcomes than any others we could realize, ought not to be brought about because they cannot be achieved without violating people's fundamental rights. In other words, we may believe in *side-constraints* on permissible harm (Nozick 1974: 28-33). Or we may believe that although some outcome would be morally best, we are not obligated to bring it about because the personal cost would be too great. In other words, we may believe in *agent-centred prerogatives* that allow us to care about our own interests out of proportion to their significance considered impartially (Scheffler 1991). There is, therefore, a logical gap between axiological longtermism and *deontic longtermism*, where the latter holds that in a wide class of decision contexts, the option one ought to choose is the option that corresponds to the best lottery over histories from t onwards, where t is some date far in the future.

Is there some way to bridge this gap? Or might it be that axiological longtermism is true while deontic longtermism is false?

Greaves and MacAskill propose to bridge this gap by appeal to a *stakes-sensitivity argument*. As we have noted, not only is humanity's total potential future greater than its near-term future history, it is *vastly* greater. As a result, it is plausible that the best ways of affecting the long run future are vastly better than the best ways of affecting the short run. Greaves and MacAskill appeal to the principle that when the axiological stakes are high, non-consequentialist constraints and prerogatives tend to be insignificant in comparison, so that what one ought to do is simply whichever option is best. Let's call this the *Stakes Principle*.

The Stakes Principle is plausible on its face. It is natural to think that in order to decide on the right action in any given case, we need to weigh up the different competing considerations that apply in that case. Promoting the good matters, but so does respect for individual rights, and

loyalty to one's self and one's nearest and dearest. Balancing these competing considerations means that if any one among them is sufficiently weighty in comparison with the others, it will carry the day, just as it is always possible to tip a pair of scales by adding sufficient weight to one side. Respect for individual rights is important, but if the good that could be achieved by violating some right is great enough, that is what we should do. So says the Stakes Principle, and so say most of us.

In this paper, I will argue that the Stakes Principle is nonetheless false. Furthermore, I will present a *prima facie* plausible view about the relationship between value and obligations in the context of the Non-Identity Problem that suggests that axiological longtermism is true but deontic longtermism is false, even granting that the value at stake over the very long run is astronomical in comparison with the value at stake within the near-term. The overall plausibility of this view deserves further scrutiny, but indicates the existence of a significant obstacle to bridging the gap between consequentialist and deontological perspectives on the moral significance of the long run future.

2.

In assessing the Stakes Principle, we should begin by noting that it isn't altogether transparent. The principle says that when the axiological stakes are high, non-consequentialist constraints and prerogatives tend to be insignificant in comparison. What does it mean to say that 'the axiological stakes are high'?

We could interpret this phrase so that the Stakes Principle is tautologous. In other words, we could interpret 'the axiological stakes are high' to mean that the value at stake is such that non-consequentialist constraints and prerogatives tend to be insignificant in comparison. This

would make the principle unassailable, but at the cost of leaving its application in any given instance undetermined.

Alternatively, we could interpret 'the axiological stakes are high' to mean simply that there is a lot of value at stake. Obviously, this is still a vague requirement. What constitutes 'a lot'? Nonetheless, I think we are able to construct a plausible case that the Stakes Principle, so interpreted, ought to be rejected. At least, we can show that the Stakes Principle conflicts with a view about aggregation that is widely endorsed by contemporary non-consequentialists.

Consider one of the most widely discussed thought experiments in recent moral philosophy (Scanlon 1998: 235):

Transmitter Room

Jones is trapped under electrical equipment in the transmitter room of a TV station, receiving painful electrical shocks. To rescue him, we must interrupt the broadcast of a World Cup match that otherwise will not end for one hour.

Many have the intuition that we ought to save Jones, no matter how many people will be inconvenienced as a result of having the broadcast interrupted. A wide range of theories have been developed in order to capture the intuition that very significant goods have lexical priority with respect to trivial goods: i.e., we ought to provide a single very significant benefit to one person rather than provide trivial benefits to arbitrarily large groups of people (see, *inter alia*, Kamm 1993, Scanlon 1998, Voorhoeve 2014).

What is the significance of *Transmitter Room* for the Stakes Principle? It is natural to assume that the enjoyment that people derive from watching the World Cup match is good. Assume that moral goodness satisfies the *Archimedean property*: for any two goods, *a* and *b*, there is

a natural number, n , such that n units of a is more valuable than one unit of b . In that case, the goodness at stake in *Transmitter Room* can be made arbitrarily high by increasing the viewership. Nonetheless, the intuition is that we ought to save Jones, no matter how many people will be inconvenienced as a result of having the broadcast interrupted. If we take this thought seriously, then it seems we ought to reject the Stakes Principle.

Someone might object that *Transmitter Room* is not really a challenge to the Stakes Principle. They may think that it does not pit characteristically non-consequentialist moral considerations against characteristically consequentialist considerations. They may insist that the *Transmitter Room* case is instead a challenge to the assumption that goodness satisfies the Archimedean property. That's because someone may believe that no matter how many people would be inconvenienced, the outcome in which Jones is saved and those people are annoyed at having their enjoyment interrupted is a *better* outcome than that in which Jones is left to suffer and the viewers are able to enjoy the match uninterrupted. In other words, very significant benefits are lexically better than trivial benefits (Temkin 1996, 2012). The choice to save Jones is therefore not a matter of privileging characteristically non-consequentialist moral concerns over characteristically consequentialist moral concerns, regardless of how much good is at stake, but simply a matter of privileging better outcomes over worse outcomes. At the same time, the *Transmitter Room* case shows that goodness does not satisfy the Archimedean property.

However, a well-known spectrum argument challenges the claim that saving Jones yields the better outcome (Norcross 1997; Temkin 1996, 2012). Imagine that if we left Jones to suffer, we could save some large number of people experiencing pain that is only slightly less intense. Surely, it is better if the pain of the many is alleviated. Although it is slightly less intense, it is suffered by many more people. However, we can then go on imagining larger populations of people who would experience slightly less bad personal misfortunes, in each case judging that

helping the greater number is better. We can continue to decrement the badness of what is experienced while increasing the number of people who experience it until we come upon a suitably large group of people who may be inconvenienced by having the broadcast of an enjoyable World Cup match interrupted. Because in every pairwise comparison it is deemed better if the greater numberer who would experience the less serious personal bad are helped and because it is compelling to suppose that *__better than __* is a transitive relation (*pace* Rachels 1998, Temkin 1996, 2012), it follows that preventing so-and-so-many people from being inconvenienced by having the broadcast interrupted represents a better outcome than saving Jones. In demonstrating that sufficiently many trivial goods can outweigh something of very great importance, this argument also supports the more general assumption that goodness satisfies the Archimedean property.

Why can't a similar argument be used to show that we ought to allow so-and-so-many people to continue enjoying the World Cup match rather than rescuing Jones? After all, in a choice between saving Jones or saving some suitably large group of people experiencing pain that is only slightly less intense, we ought surely to help the people in the suitably large group. And we seem to be able to go on imagining larger populations of people who would experience slightly less bad personal misfortunes, such that in each case it seems obvious that we ought to help the greater number, thereby constructing a spectrum of cases just like that described in the previous paragraph. If, in each pairwise comparison, we ought to help the greater number, how do we avoid arriving at the analogous conclusion that we ought to leave Jones to suffer rather than inconvenience the many viewers of the World Cup game?

The key difference is that whereas it seems very hard to deny that *__ better than_* is a transitive relation, there are compelling reasons for non-consequentialists to reject the principle that *ought to do _ rather than _* is transitive (Friedman 2009; Kamm 1985; Willenkin 2012). Consider

the following sequence of cases discussed by Frick (2014: 119), based on Kamm (2007: 26). You ought to divert a train away from five people toward one person whom it will kill rather than let it kill the five. Furthermore, you ought to let the train kill the five rather than push someone into its path who will stop its motion but incur injuries so severe that they leave her with a life that is barely worth living. However, it is not the case that you ought to divert a train away from five people toward one person whom it will kill rather than push that person into its path in a way that will stop its motion but leave her with injuries such that the rest of her life is only just worth living, because the latter is better for everyone and worse for none. We thus have a counter-example to the transitivity of *ought to do _ rather than _*.

In sum, there is a strong case for thinking that when the number of people watching is sufficiently great, the impartially better outcome in *Transmitter Room* is that in which Jones is left to suffer. Furthermore, the amount of value at stake can be made arbitrarily great by making the number of viewers arbitrarily great. Nonetheless, there is a plausible case for thinking that you ought to save Jones no matter how many people are watching. Therefore, there is a plausible case for rejecting the Stakes Principle.

Note, moreover, that the claim about the *Transmitter Room* case on which we have focused so far is a strong one: that you ought to save Jones, no matter how many are watching. This is the standard claim about this sort of case favoured by contemporary non-consequentialists. But it is stronger than is needed to challenge the Stakes Principle. All we need is that it is *permissible* to save Jones, no matter how many people are watching. Even those who doubt that you are *required* to save Jones regardless of the numbers may find this verdict compelling.

3.

Transmitter Room illustrates one way in which the Stakes Principle may plausibly be thought to fail: namely, when we are weighing very significant benefits against trivial benefits. But it may be thought that this is not what is at issue when it comes to making a choice between promoting a flourishing long-term future for humanity versus attending to problems confined to the near term.

Greaves and MacAskill are actually not persuaded of this. They suggest that the question of how to weigh very significant benefits against trivial benefits may be of issue when it comes to the case for long-termism, once we consider how the view that very significant goods have lexical priority with respect to trivial goods may be interpreted in respect of *ex ante* goods: i.e., lotteries over benefits.

In particular, consider the view that a lottery that assigns a very, very low probability to a very significant good yields a trivial *ex ante* benefit, while a lottery that assigns probability 1 to some very significant good yields a very significant *ex ante* benefit, taken in conjunction with the principle that very significant *ex ante* goods have lexical priority with respect to trivial *ex ante* goods. According to Greaves and MacAskill, “the ways in which typical long-termist interventions deliver high expected value is via a very small probability of a significant benefit to each of an enormously large number of (possible) future persons.” (18-19) The thought here is that a typical longtermist intervention will aim to raise the average welfare over the long-run, but is very unlikely to succeed (though still recommended in expected value terms). Therefore, someone holding a view of the kind we have just described may believe that concern for the long-term future should typically be set aside, because there are typically identified individuals living in the present moment to whom we can provide very significant benefits with (near) certainty, and

benefits of this kind have lexical priority with respect to extremely small chances of very significant goods.

One complication here is the issue of *non-identity*. We may expect any action with the potential to impact on the long-term future of humanity to change the identities of the people who exist in the long-term future. For this reason, we may deny that long-termist interventions yield a very small probability of a very significant *benefit* to each of an enormously large number of possible future persons. Since any person who exists if the good effects of that intervention are actualised would not have been worse off had the intervention not been undertaken, we may deny that they are benefitted. We may think that in order to be benefitted, you must be made better off than you would otherwise have been.

Here is a second reason to think that we have not hit on the crux of the debate. No one should take seriously the view that in a choice between providing someone with a very significant benefit that they will receive with certainty and providing each person in some larger group with a very, very small probability of receiving some comparably significant good, we ought to help the former person no matter how many people are in the other group. This view would imply that we ought to prioritise saving a single identified life over *any number* of statistical lives. Thus, in a choice between saving the life of some particular person who is dying of a disease or vaccinating the entire population so that they become immune to the disease, this view tells us that we ought to save the single individual even if the vaccination campaign would save millions, or even billions, or even trillions, etc. Whereas the view that we ought to prioritise saving *an* identified life over *a* statistical life has some plausibility, the view that the saving an identified life has lexical priority is extremely implausible (Frick 2015a, 2015b).

I conclude that the question of how to weigh very significant benefits against trivial benefits is not where the action is when it comes to the question of whether there is a sound

argument from axiological longtermism to deontic longtermism, given suitable additional premises. The significance of the *Transmitter Room* case is merely in providing us with a very clear challenge to the Stakes Principle, indicating that if there is a sound argument of that kind, it will not include this principle as a premise.

4.

Instead of the question of how to weigh benefits of very different magnitudes, I think the central issue is the *Non-Identity Problem*. By this, I have in mind the question of how to evaluate same-number cases in which our choice determines both the identity and the welfare level of some future person(s), such as

Conception or Delay

Hiroko plans to conceive a child. She learns that she has contracted a subclinical infection such that if she conceives a child now, then her child will be born blind. If she delays conception by a year, the infection will clear, and any child she conceives then will be born sighted.

Let us assume, as most people believe, that blindness lowers a person's welfare. (If you believe instead that blindness typically yields benefits that offset whatever costs are associated with the absence of sight and therefore does not significantly decrease a person's quality of life (Barnes 2016: 94-96), feel free to substitute some other property into the description of *Conception or Delay* that you believe would satisfy this criterion when thinking about this case.) Granting the assumption, let us ask: what should Hiroko do? Many people think she morally ought to wait. However, her child is not made better off by this choice, since the child who will be born a year hence is not the same child who would have been born blind had Hiroko decided not to wait.

Why should thinking about what Hiroko ought to do be of interest to us? In building their case for axiological longtermism, Greaves and MacAskill note that there are some views on which we can achieve astronomical value by making the future population much greater than it would otherwise have been (Beckstead 2013; Bostrom 2008, 2013). Greaves and MacAskill want to bracket the controversial question of whether adding lives worth living to the population is intrinsically good, and so their discussion focuses instead “on the prospects for generating large amounts of expected value by improvements to the expected *average well-being* of future people,” (7) while noting that the “interventions in question, despite (perhaps) not significantly altering future population size, nonetheless change which (far-)future people are brought into existence.” (19) In other words, we are to focus on cases in which our choice has the potential to alter the identities and raise the welfare-level of future people, while having a morally trivial impact on their number.

Apart from the simplification that they are assumed to involve exactly the same number of people and do not involve uncertainty about the final outcome, cases illustrative of the Non-Identity Problem are cases of exactly this kind. And here we are on much safer ground in terms of the core axiological assumptions required to run the argument, since many people who reject the view that one outcome is better than another if it contains additional lives that are worth living are nonetheless attracted to Parfit’s *Same-Number Quality Claim*, according to which, between two populations containing the same number of people (but not necessarily the same people), if one outcome involves a higher average level of well-being, then it is better, all else being equal.

Admittedly, there is one important aspect of the choice between short-term focused and long-term focused interventions missing from cases like *Conception or Delay*: they do not involve a choice between changing the identity and the welfare level of some future person(s) versus

benefiting some other already existing person (except insofar as Hiroko is benefited by conceiving now as opposed to later). I intend to set aside this point, as I want to focus our attention on the plausibility of the following (initially counter-intuitive) view about *Conception or Delay*: a view according to which, although it would be better if Hiroko waited and conceived a child one year from now, it is not obligatory. If it is not in general obligatory in cases illustrative of the Non-Identity Problem to choose the act that leads to greater well-being for some future person(s) when there is no one else who could be said to have a claim on us to do something else, then it surely won't be obligatory when such a person is added to the mix. Therefore, we can ignore the existence of such persons for the time being.

5.

Why do I say that it is plausible to suppose that although it would be better for Hiroko to wait and conceive a child in one year, it is not obligatory? I won't say anything to justify the verdict that it would be better to delay conception, as I don't think that is where the point of contention lies. Instead I will address the claim that delaying conception is not obligatory. However, it is worth emphasizing that the view we are asked to accept here does include the claim that delaying conception is better. It does not imply that there is no moral reason to delay conception. It merely insists that such reasons do not ground an obligation for Hiroko to delay. A view of this kind is much easier to believe than the view on which it would not be morally better if Hiroko had a child with a higher quality of life. While the latter is extremely counter-intuitive, the former is arguably only mildly so.

Why should we think it is nonetheless believable? Firstly, we can present considerations that seem to shift the burden of proof (Boonin 2014: 198-205). Consider the following cases:

*Conception or Delay**:

Hiroko learns that she has contracted a subclinical infection such that if she conceives a child now, then her child will be born blind, but otherwise healthy. If she delays conception, she will not have a child.

*Conception or Delay***:

Hiroko learns that if she conceives a child now, then her child will be born sighted and otherwise healthy. If she delays conception, she will not have a child.

Intuitively, in *Conception or Delay**, it is permissible for Hiroko to choose to give birth to a child who will be born blind rather than having no child. In *Conception or Delay***, intuitively, Hiroko is permitted to choose to have no child rather than having a sighted child now. So it is permissible for Hiroko to choose to have a child who will be born blind rather than no child, and it is permissible for her to choose to have no child rather than a child that would be sighted and otherwise healthy. How, then, could it be impermissible for Hiroko to choose to have a child who will be born blind rather than a child that would be sighted and otherwise healthy, as in the original *Conception or Delay* case?

It may seem that this argument presupposes that *permissible to do _ rather than _* is a transitive relation. This would be unfortunate, as there exist compelling counter-examples to that assumption (Kamm 1985). For example, it is permissible to try to save a life at great personal risk rather than to keep a promise to meet a friend for lunch, and it is permissible to spend the afternoon playing video games rather than try to save a life at great personal risk, but it is not permissible to spend the afternoon playing video games rather than keep a promise to meet a friend for lunch.

However, the point of the previous paragraph was not to *prove* that Hiroko is permitted to conceive without delay. The aim was merely to shift the burden of proof. The assumption is that

cases in which *permissible to do _ rather than _* relates φ to χ in a pairwise comparison of these options and χ to ψ in a pairwise comparison of these options, but not φ to ψ in a pairwise comparison of these options are surprising and stand in need of explanation. (All too briefly: what explains the failure of transitivity in the case described immediately above is that justifying and requiring strength are not always equal when it comes to moral reasons. See section 7 for further discussion.) Anyone who grants that it is permissible for Hiroko to choose to have a child who will be born blind rather than no child and permissible for her to choose to have no child rather than a child that would be sighted and otherwise healthy but who insists that Hiroko may not choose a child that will be born blind rather than a similar child that will be born sighted owes us an explanation of this surprising fact. However, no satisfactory explanation is in the offing.

We can begin by addressing what is perhaps the most obvious explanation that might occur to us: namely, that there is a morally relevant difference between *Conception or Delay* and *Conception or Delay** in that in the latter a prohibition on choosing to have a child who will be born blind would be excessively demanding, since this would require Hiroko to have no child at all, whereas in the former no similar complaint can be raised against a prohibition on choosing the child who will be born blind, because there is another option by which Hiroko can become a mother.

The problem with this reply is that we need not imagine that having no child would be costly for Hiroko. We may imagine that in every choice she faces, she would be equally well off regardless of what choice she makes. Thus, in *Conception or Delay** we can imagine that she is equally drawn to becoming the mother of a child who will be born blind and foregoing motherhood in order to devote all her time to some other project, and neither life would be better or worse for her overall, because she would find meaning and satisfaction in either. From the point of view of her own interests, there is nothing to choose between the two options. Even in

that case, it does not seem plausible to suppose that she is not permitted to become the mother of a child who will be born blind rather than remaining childless.

Note also that an axiological explanation seems unpromising. We may agree that it would be better if Hiroko chose to wait in *Conception or Delay*. However, we are not generally obligated to maximize the good. Furthermore, in *Conception or Delay***, it is plausible that the outcome would be better from the moral point of view if Hiroko were to have a child who will be healthy and happy rather than having no child at all (Broome 2005; Huemer 2008). Therefore, if we wish to explain why Hiroko is required to have a sighted child rather than a blind child by appeal to the fact that the former would be better, we face the problem of explaining why she is not obligated to have a child who will be healthy and happy, rather than having no child at all.

According to Harman (2004), conceiving a child who will be born blind harms the child who will be born, because a sufficient condition for one person to harm another is that they cause that person to exist in a state in which they suffer pain, early death, bodily damage, or deformity. Therefore, we are supposed to be able to explain why Hiroko should wait in *Conception or Delay* by appeal to the claim that she will otherwise cause significant, avoidable harm to her child. A *prima facie* problem for this view is that, on Harman's conception of harm, it is also true in *Conception or Delay** that refusal to delay will cause significant harm to Hiroko's child. Harman's response is that the key difference between *Conception or Delay* and *Conception or Delay** is that in *Conception or Delay* there exists an alternative that would involve parallel benefits, without parallel harm, whereas there is no such alternative in *Conception or Delay**. (Just as we count the bad events that will arise in a child's life as harms caused by her conception, so we count the good events that will arise in her life as benefits caused by her conception.) But this will not do, because it is not generally the case that if one faces two alternatives, φ and ψ , such that ψ yields parallel benefits without parallel harm, then choice of ψ is obligatory. For example, if a doctor can perform a

life-saving operation on either of two patients whereas only one of these patients would have a painful recovery while the other would recover painlessly, it is not obligatory that the doctor saves the latter.

There is much more that could be said on this topic (see Boonin 2014). Nonetheless, I hope I have said enough at this point to allow the reader to see why we may think it reasonable to deny that Hiroko acts wrongly in *Conception or Delay* if she conceives without delay, although delaying would bring about the better outcome.

6.

Let us now return to the question of what significance this denial might carry in evaluating the possibility of a sound argument from axiological longtermism to deontic longtermism. It seems natural to think that if the case for axiological longtermism rests ultimately on the value of increasing the average welfare level of future generations without altering their number in any morally significant way, then a sound argument from axiological to deontic longtermism exists only if it is obligatory to choose the population with higher wellbeing in cases illustrate of the Non-Identity Problem, such as *Conception or Delay*. Therefore, if it is not obligatory to choose in this way in such cases, there exists no sound argument from axiological to deontic longtermism.

However, we may query the supposition that a sound argument from axiological to deontic longtermism exists only if it is obligatory to choose the population with higher wellbeing in cases like *Conception or Delay*. This may be thought to ignore the key observation associated with the stakes-sensitivity argument put forward by MacAskill and Greaves: namely, that so much is at stake when it comes to the long-term future of humanity. In *Conception or Delay*, it would be better if Hiroko delayed conception. But it would not be vastly better. At least, it would not be vastly better in the same way that it would be vastly better if the average welfare level were

(non-negligibly) higher in every generation existing between the year 2312 and the heat death of the universe. Because we are not dealing with vast differences in value when we consider *Conception or Delay*, what we think about this case need not have the kind of significance it was supposed to have in the previous paragraph.

However, it is reasonably straightforward to bridge this gap. If Hiroko is permitted to conceive without delay in *Conception or Delay*, then she would presumably be permitted to make this choice over and over, if she were somehow to face it again and again while the choice remained similar in all relevant respects on each occasion. And it stands to reason that it would be permissible for her to resolve to make the decision over and over, if she were expecting somehow to face it again and again. But in a choice between repeatedly conceiving without delay and repeatedly delaying conception, we may arrive at a choice between options that differ arbitrarily in value. Plausibly, the difference in value between these plans is a positive linear function of the number of iterations. The difference in value can therefore be made arbitrarily great by making the number of iterations arbitrarily great. Nonetheless, we have seen that there is a strong case for inferring that Hiroko is permitted to choose the suboptimal plan if she is permitted to choose to conceive without delay in *Conception or Delay*. In light of this, I think we should conclude that the fact that we are not dealing with vast differences in value in *Conception or Delay* does not undercut the assumption that a sound argument from axiological to deontic longtermism exists only if it is obligatory to choose to conceive the child whose lifetime wellbeing will be higher in *Conception or Delay*.

Note also that I say 'only if' and not 'if and only if.' In arguing against the existence of a sound argument from axiological to deontic longtermism by denying the existence of an obligation for Hiroko to maximize the welfare level of her child in *Conception or Delay*, I have made my task harder than it need be. You could think that *Conception or Delay* is not the right sort of test

case for thinking about the relationship between axiological and deontic longtermism, not because this is unfavourable to those who believe in a sound argument from the former to the latter, but because it stacks the deck in their favour.

Here is why. *Conception or Delay* is like so many other cases chosen to illustrate the Non-Identity Problem in that it not only involves a choice between non-identical equinumerous populations with different welfare levels, but a choice between non-identical equinumerous populations with different welfare levels wherein the population with the lower welfare level falls below some relevant threshold of *normality*. Thus, blindness is a disability, a failure of the eyes and/or visual cortical areas to fulfil their biologically defined 'proper function'. Someone who is worse off than others because she is blind is therefore worse off by virtue of the absence of a good that defines what we take to be the human biological norm. Where a potential person stands with respect to some implicit threshold of normality seems to make a significant difference to our intuitions about cases of this kind. Whereas people's intuitions tend to favour the view that someone like Hiroko morally ought to wait in order to conceive a child who will not be worse off than others due to the failure of her eyes and/or visual cortical areas to fulfil their 'proper function,' we do not have similar intuitions about otherwise comparable cases wherein delaying conception would allow Hiroko to conceive a child whose lifetime wellbeing would be higher due to the fact that delaying would allow her to conceive a child with supranormal sensory abilities, whereas conceiving now would allow her to conceive a child with merely normal abilities (Lillehammer 2005). Similarly, while many have the intuition that Hiroko ought to delay conception if she would otherwise conceive a child with severe cognitive impairments whose mental abilities are no different from those of a horse, no one thinks that Hiroko morally ought to choose to conceive a child with normal mental abilities rather than breeding a horse, because whereas a child whose mental abilities are no different from those of a horse is considered

abnormal, a horse with those mental abilities is entirely normal relative to the species to which it belongs (Boonin 2014: 205-209).

What is the significance of this? On the one hand, it helps to make the case that the burden of proof falls on those who insist that Hiroko is obligated to delay, by providing a reason for us to be suspicious of our intuitions. I am deeply sceptical that we can find any plausible standard of normality on which that standard has the kind of moral significance it must have if these intuitions are to be trusted. But establishing this is beyond the scope of this paper. Furthermore, it is not totally clear in what way we should come to distrust our intuitions if we granted the conclusion. It is not clear whether we should come to distrust the intuition that Hiroko morally ought to delay conception in order to conceive a child with normal abilities or instead the intuition that Hiroko is not morally required to delay conception in order to conceive a child with supranormal abilities, as suggested by Savulescu (2001).

In any case, when we consider “prospects for generating large amounts of expected value by improvements to the expected *average well-being* of future people,” these improvements are unlikely to involve replacing people whose well-being would fall below some relevant threshold of normality with different people whose well-being is at or above that threshold. More often than not, we may expect, they involve replacing people whose well-being would be at or above that threshold with people whose well-being is even higher. After all, people who live in the future will, on average, be much richer than people who live today, and advances in medicine will allow them to conquer diseases that afflict current generations. Thus, when we consider “prospects for generating large amounts of expected value by improvements to the expected *average well-being* of future people,” we are not simply dealing with cases that invoke the Non-Identity Problem. We are dealing with cases that invoke those instances of the Non-Identity

Problem wherein the verdict that we are morally required to choose the population with higher well-being is least compelling.

6.

My approach so far in this paper has been driven by intuitions about cases. I have said very little about what the deep structure of morality must be like in order to support these intuitions. But something ought to be said about this, because it is surprising that the *Stakes Principle* fails for the sort of choices we're considering. How can it be that with so much value at stake, axiological considerations fail to ground an obligation? Presumably, the explanation will be very different from the sort of theoretical principles that are invoked to justify our intuitions about cases like *Transmitter Room*, which typically appeal to a model of competing claims and constraints on reasonableness among claimants. So what is it?

In order to address this question, we need to reflect on the relationship between moral value, moral reasons, moral oughts, and moral obligations. There is a straightforward picture of the relationship between these elements of moral assessment on which the *Stakes Principle* arguably draws. We assume

- (i) If some outcome is morally valuable, then there is a moral reason to bring it about whose strength is proportional to its value.
- (ii) If that moral reason is not opposed by sufficiently strong reasons that recommend some other action, then the agent morally ought to bring about that outcome.
- (iii) If the agent morally ought to bring about that outcome, she is obligated to do so.

This is the sort of picture I think we ought to resist.

Tooley (1998) has suggested that although adding people with lives worth living to the population makes the outcome morally better, this does not entail that there is a moral reason for making happy people. In other words, Tooley denies (i). But Tooley's reason for rejecting (i) seems to be, in large part, that he does not think we can reject (ii) or (iii), and he wants to avoid the conclusion that we are obligated to make happy people. He writes: "In the present context, a reason is a moral reason, and so to say that something provides one with a reason for performing a given action is to say that the relevant feature is a right-making property of actions, and thus that actions having that feature are prima facie obligatory." (117) But it is arguably more plausible to reject (ii) or (iii), and so I will not say more about Tooley's position here (see Frick 2014: 65-66 for further discussion).

Dancy (2004) argues for the existence of 'enticing reasons,' where p is an enticing reason for S to φ only if p is a reason to φ and it is not the case that: if there is no reason for S not to φ , then S ought to φ in light of the reason for φ 'ing provided by p . Enticing reasons are not in the business of telling us what we ought to do, so much as rendering intelligible some choice, by highlighting attractive features of its object. They characteristically have to do with what would be fun, amusing, or exciting, the thought being that reasons of this kind are too flimsy to ground an 'ought,' but are reasons nonetheless. Dancy writes in a way that suggests that moral reasons are never enticing, but this may be doubted by someone who believes in enticing reasons. For example, we could believe that there are moral reasons that support doing small favours or providing trivial benefits to others, but that these do not rise to the level of moral 'oughts' (compare Kagan 1989: 243-44).

Alternatively, we may prefer to deny the link between ought and obligation. Thus Broome (2004) rejects the existence of Dancy's enticing reasons, noting that it seems alright to say, for example, that someone ought to try a mangosteen if they have not done so, although the only

reason for trying them is that they are delicious. Instead, Broome suggests that we can better capture the sort of phenomenon to which Dancy is trying to draw our attention by rejecting the assumption that you are obliged to do what you ought to do. While it seems okay to say that you ought to try a mangosteen if you have never done so before, saying that you are obliged to try one seems hyperbolic. It may well be that this has something to do with prudential reasons in particular: there is arguably something forced about speaking of ‘obligations of prudential normativity,’ regardless of how much is at stake. However, someone might claim that there are also some moral reasons that ground oughts but not obligations (compare Kagan 1989: 64-70).

We should also note Gert’s (2003) distinction between the *justifying* and *requiring* strength of reasons. On this view, there are two roles for reasons to play. Firstly, they can render permissible responses that would otherwise be impermissible. The capacity of a reason to perform this role provides a measure of its ‘justifying strength.’ Secondly, they can render impermissible responses that would otherwise be permissible. The capacity of a reason to perform this role provides a measure of its ‘requiring strength.’

Gert’s key insight is that justifying strength and requiring strength need not be equal. A reason may have significant justifying strength but weak requiring strength. We noted earlier that it is permissible to try to save a life at great personal risk rather than keep a promise to meet a friend, but also that it is permissible to stay home playing video games rather than try to save a life at great personal risk, although it is not permissible to stay home playing video games rather than keep a promise to meet a friend. We can explain this by appeal to the idea that the fact that an action would save a life but at great personal risk is a reason with very high justifying strength, but limited requiring strength. In particular, it has less requiring strength than promise keeping. It can therefore justify breaking a promise, but cannot require you to give up an afternoon of

trivial enjoyment, whereas having made a promise can require you to give up an afternoon of trivial enjoyment.

Gert (2000) notes that not only can justifying and requiring strength come apart, there also seem to be purely justifying moral reasons: i.e., reasons that have positive justifying strength but *no* requiring strength. Self-interest seems to feature in this way in many moral decisions. Appeal to one's own self-interest can justify failure to perform some action that would be beneficial to others, although we are not inclined to think the agent would be morally required to act in her own self-interest were there nothing else worthwhile she could be doing. Rarely, if ever, is prudence morally obligatory. Supposing that there are purely justifying reasons, either (ii) or (iii) is false, though we cannot say which.

9.

Let us now put our taxonomic observations to use. The view we are to consider is that additional lives worth living are intrinsically good and better additional lives are morally better, whereas the fact that some act would bring into existence a person with a life worth living is a purely justifying reason whose strength is proportional to the goodness of her life. This reason has zero requiring strength.

Call this the *Non-Requiring View*. If the Non-Requiring View is correct, this would explain why it would be better for Hiroko to wait in *Conception or Delay*, and why it makes sense to say that she has greater moral reason to wait, although she has no obligation to wait. Depending on whether we reject (ii) or (iii) in acknowledging the existence of purely justifying reasons, we might even be able to say that Hiroko ought to wait, but is not required to do so.

The Non-Requiring View also allows us to capture the widely held intuition that a couple is not required to conceive a child, even if they know any child they conceive would have a happy

life. The fact that they could conceive a happy child is not a reason with requiring strength. But the Non-Requiring View also straightforwardly explains the intuition that the couple is permitted to conceive a child even if they know that this will involve added burdens for themselves, for their already existing children, or for the environment. That is because it allows us to say that the value of an additional happy life has considerable justifying strength.

Many people also have the intuition that if a couple knows that any child they will conceive would have a life that is full of suffering, then they are under an obligation not to bring such a child into existence. We therefore cannot say that the fact that an act would bring into existence a life that is not worth living is a reason against doing so with zero requiring strength. Unsurprisingly, if we want to capture these intuitions, we have to postulate a fundamental asymmetry between additional good and bad lives, counting the former as sources of merely justifying reasons, while attributing non-zero requiring strength to the latter. Call this the *Non-Requiring View+*.

The Non-Requiring View+ is similar to a view proposed by McMahan (2013). McMahan defines an intrinsically good event that would happen in the life of some potential person if she is created and who would otherwise not exist as a *non-comparative existential benefit* conferred by her conception. He considers the suggestion that non-comparative existential benefits have no *reason-giving weight*, in that we do not have reasons to bring them about, but they nonetheless have *cancelling weight*, in that the existential benefits that a potential person will enjoy can outweigh the reasons against bringing her into existence that obtain in virtue of the non-comparative existential harms she will suffer, which do have reason-giving weight.

The Non-Requiring View+ is obviously very similar to the view McMahan discusses. However, it differs from that view in a number of respects. Firstly, the Non-Requiring View+ avoids the conceptual awkwardness of postulating considerations that have no reason-giving

weight but have cancelling weight. It is not at all clear how to make sense of the idea that something which is not a reason in favour of some action can outweigh something which is a reason against performing that act. Surely *outweighing*, in this sense, is a relation between reasons. We are forced to deny this on the view proposed by McMahan, but not on the Non-Requiring View+.

We are also not driven to say that the thing that ultimately leads McMahan to reject the view he considers: namely, that in a case life *Conception or Delay*, there is no reason to delay, because non-comparative existential benefits are not reason-giving, and hence it cannot be the case that we have greater reason to choose options that bring about greater non-comparative existential benefits. On the Non-Requiring View+, we have greater reason to bring about greater non-comparative existential benefits, but these reasons are purely justifying reasons.

Last but not least, on the view described by McMahan, it appears that the cancelling weight of non-comparative existential benefits is purely intrapersonal. McMahan describes cancelling weight as “the weight that noncomparative benefits have in canceling the noncomparative harms suffered *by the same person*” (21 - my emphasis). By contrast, the Non-Requiring View+ is defined with respect to whole lives and entails that creating lives worth living has justifying strength with respect to bringing into existence lives that are not worth living.

Importantly, this means that we can justify allowing the human race to go on existing, whereas this is hard to account for within the terms of McMahan’s view. After all, we can be as good as certain that the continued existence of the human species will lead to the existence of many lives that are not worth living, alongside the far greater number of lives that are well worth living. Every so often, someone will be born afflicted by some terrible genetic disease in light of which she is forced to endure a life full of suffering. Suppose we think that there is no reason to

bring into existence lives that are well worth living but an obligation to prevent the existence of lives that are not worth living, and the goods within lives of the former kind do not have cancelling weight with respect to the bads in lives of the latter kind. Then we are forced to conclude that we ought to bring about the end of all sentient life - or at least to vastly curtail the future population, insofar as there are non-welfarist reasons related to species-conservation that can justify us in keeping it going (see Parfit 1984: 409-11, Frick 2017 for discussion).

9.

In Act II of Checkhov's *Three Sisters*, the lieutenant colonel Aleksandr Vershinin encourages the other characters to discuss what they imagine life will be like in the future. His own view is that the present exists for the sake of the glorious hereafter. He tells the lieutenant, Tuzenbak: "In two or three hundred, perhaps in a thousand years - the time does not matter - a new, happy life will come. We shall have no share in that life, of course, but we're living for it, we're working, well, yes, and suffering for it, we're creating it - and that alone is the purpose of our existence, and is our happiness, if you like."

There is something at once inspiring and chilling about this vision of our place in history. I hope this paper has managed to capture something of that duality. I grant axiological longtermism, but I have argued against the Stakes Principle and presented reasons to doubt that there is any sound argument from axiological longtermism to deontic longtermism, at least insofar as we are concerned with ways of improving the value of the future of the kind that are focal in Greaves and MacAskill's presentation. The future need not have priority over the present, therefore. It need not be the purpose of our existence. Nonetheless working for it - and even suffering for it - may be in the service of the highest good.

Bibliography

- Archer, David, Michael Eby, Victor Brovkin, Andy Ridgwell, Long Cao, Uwe Micolajewicz, Ken Caldeira, Katsumi Matsumoto, Guy Monhoven, Alvaro Montenegro, and Kathy Tokos (2009) Atmospheric lifetime of fossil fuel carbon dioxide. *Annual Review of Earth Planetary Science* 37, 117-34.
- Barnes, Elizabeth (2016) *The minority body: a theory of disability*. Oxford: Oxford University Press.
- Beckstead, Nick (2013) On the overwhelming importance of shaping the far future. PhD thesis: Department of Philosophy, Rutgers University.
- Boonin, David (2014) *The Non-Identity Problem and the ethics of future generations*. Oxford: Oxford University Press.
- Bostrom, Nick (2008) Astronomical waste: the opportunity cost of delayed technological development. *Utilitas* 15, 308-14.
- (2013) Existential risk prevention as a global priority. *Global Policy* 4, 15-31.
- Broome, John (1991) *Weighing goods*. Oxford: Blackwell.
- (2004) Reasons. In Wallace, Pettit, Scheffler, and Smith, eds. *Reason and value: themes from the moral philosophy of Joseph Raz*, 28-55. Oxford: Oxford University Press.
- (2005) Should we value population? *Journal of Political Philosophy* 13, 399-413.
- Bykvist, Krister (2007) The benefits of coming into existence. *Philosophical Studies* 135, 335-362.
- Dancy, Jonathan (2004) Enticing reasons. In Wallace, Pettit, Scheffler, and Smith, eds. *Reason and the moral philosophy of Joseph Raz*, 91-118. Oxford: Oxford University Press.
- The Economist (2013) Towards the end of poverty. 01/06/2013

- Frick, Johann (2014) 'Making people happy, not making happy people': a defense of the asymmetry intuition in population ethics. PhD Thesis: Department of Philosophy, Harvard University.
- (2015a) Contractualism and social risk. *Philosophy & Public Affairs* 43, 175-223.
- (2015b) Treatment versus prevention in the fight against HIV/AIDS and the problem of identified versus statistical lives. In Cohen, Daniels, and Eyal eds. *Identified versus statistical lives: an interdisciplinary perspective*, 182-202. Oxford: Oxford University Press.
- (2017) On the survival of humanity. *Canadian Journal of Philosophy* 47, 344-367.
- Friedman, Alex (2009) Intransitive ethics. *Journal of Moral Philosophy* 6, 277-297
- Gert, Joshua (2000) Practical rationality, morality, and purely justificatory reasons. *American Philosophical Quarterly* 37, 227-43.
- (2003) Requiring and justifying: two dimensions of normative strength. *Erkenntnis* 59, 5-36.
- Greaves, Hilary and MacAskill, William (2019) The case for longtermism. Unpublished manuscript.
- Harman, Elizabeth (2004) Can we harm and benefit in creating? *Philosophical Perspectives* 18, 89-113.
- Huemer, Michael (2008) In defence of repugnance. *Mind* 117, 899-933.
- Kagan, Shelly (1989) *The limits of morality*. Oxford: Oxford University Press.
- Kamm, Frances (1985) Supererogation and obligation. *Journal of Philosophy* 82, 118-138.
- (1993) *Morality, mortality, volume 1: death and whom to save from it*. Oxford: Oxford University Press.
- (2000) Nonconsequentialism. In LaFollette, ed. *The Blackwell guide to ethical theory*, 205-226.

- Oxford: Blackwell.
- (2007) *Intricate ethics: rights, responsibilities, and permissible harm*. Oxford: Oxford University Press.
- Lillehammer, Hallvard (2005) Benefit, disability, and the Non-Identity Problem. In Athanassoulis, ed., *Philosophical reflections on medical ethics*, 24-43. New York: Palgrave MacMillan.
- McMahan, Jeff (2013) Causing people to exist and saving people's lives. *The Journal of Ethics* 17, 5-35.
- Norcross, Alasdair (1997) Comparing harms: headaches and human lives. *Philosophy and Public Affairs* 26, 135-167.
- Nozick, Robert (1974) *Anarchy, state, and utopia*. Oxford: Blackwell.
- Parfit, Derek (1984) *Reasons and persons*. Oxford: Oxford University Press.
- Rachels, Stuart (1998) Counterexamples to the transitivity of 'better than.' *Australasian Journal of Philosophy* 76, 71-83.
- Savulescu, Julian (2001) Procreative beneficence: why we should select the best children. *Bioethics* 15, 413-426.
- Scanlon, T.M. (1998) *What we owe to each other*. Cambridge, MA: Harvard University Press.
- Scheffler, Samuel (1991) *The rejection of consequentialism, revised edition*. Oxford: Oxford University Press.
- Temkin, Larry (1996) A continuum argument for intransitivity. *Philosophy and Public Affairs* 25, 175-210.
- (2012) *Rethinking the good: moral ideals and the nature of practical reasoning*. Oxford: Oxford University Press
- Tooley, Michael (1998) Value, obligation, and the asymmetry question. *Bioethics* 12, 111-124.
- Voorhoeve, Alex (2014) How should we aggregate competing claims? *Ethics* 125, 64-87.

Willenken, Tim (2012) Deontic cycling and the structure of commonsense morality. *Ethics* 122, 545-561.