A Defence of Average Utilitarianism

MICHAEL PRESSMAN
University of Southern California

Seemingly every theory of population ethics is confronted with unpalatable implications. While various approaches to the subject have been taken, including non-consequentialist approaches, this area has been dominated by utilitarian thought. The two main approaches to population ethics have been total utilitarianism ('TU') and average utilitarianism ('AU'). According to TU, we should seek to bring about the state of affairs that maximizes the total amount of happiness. According to AU, we should seek to bring about the state of affairs that maximizes average per capita happiness. Both theories have been afflicted by seemingly strong objections, and as a result, numerous variations and hybrids have been introduced. Despite the widespread disagreement in the field, though, a near consensus has developed in rejecting AU as an absurd view. In this article, however, I will go against the grain and argue that AU is the theory of population ethics that we should endorse.

INTRODUCTION

While there have been various approaches to articulating a plausible theory of population ethics, including non-consequentialist approaches, the subject has been dominated by utilitarian thought. The two most well-known approaches to population ethics have been total utilitarianism ('TU') and average utilitarianism ('AU'). According to TU, we should seek to bring about the state of affairs that maximizes the total amount of happiness. According to AU, we should seek to bring about the state of affairs that maximizes the average amount of happiness per person. Seemingly strong objections have been made against both positions, though, and as a result numerous variations of the accounts and hybrid accounts have been introduced.

Interestingly, despite the fact that AU is such a well-known view, most philosophers who write on the topic reject it as absurd, and they focus their efforts on TU. Perhaps the main difficulty for a proponent of TU is what Derek Parfit has called the Repugnant Conclusion ('RC')[1] – an alleged *reductio* against TU – and the philosophical landscape of population ethics is largely a function of the different ways in which philosophers handle the RC obstacle. One camp of philosophers endorses TU and 'accepts' the RC, arguing that the RC does not serve as a *reductio* against TU because the RC is not in fact repugnant.[2]

[1] Derek Parfit, *Reasons and Persons* (Oxford, 1984).

[2] See Michael Huemer, 'In Defence of Repugnance', *Mind* 117 (2008), pp. 899–933; Jesper Ryberg, 'Is the Repugnant Conclusion Repugnant?', *Philosophical Articles* 25

© Cambridge University Press 2015
Utilitas Vol. 27, No. 4, December 2015
doi:10.1017/S0953820815000072

CrossMark
https://doi.org/10.1017/S0953820815000072 Published online by Cambridge University Press

Philosophers in another camp think that the RC is repugnant and that it does follow from TU, and they thus propose theories that are variations of TU that avoid implying the RC.[3] Other philosophers still – though not utilitarians – argue that the RC can be avoided by denying the transitivity of 'better than'.[4]

I think that all of these approaches are misguided, because I think that the main tenet of TU (and of most variations of TU) is wrong: it is not the case that, all else being equal, a larger population of happy people is better than a smaller one. While this conviction has surely been shared by others, I think that others prematurely abandon it when the going seemingly gets tough. In this article, I will go against the grain and argue that, if we are going to adopt a utilitarian approach to population ethics,[5] then AU is the theory we should endorse.

I will begin with a discussion of this main intuition motivating AU, and I will then explain why AU is the position we must adopt if we want to retain this intuition. I will then show why, despite its maligned status, a properly formulated AU account is a plausible account of population ethics. The bulk of the article will then be spent addressing various well-known objections to AU – including, among others, Benign Addition, Egyptology, Reverse Egyptology, and Two Hells – and I will explain why they all miss the mark.

I. PRELIMINARY POINTS

Before getting started, however, a number of preliminary points are in order. Lying beneath issues in normative ethics are various issues in metaethics, many of which cast doubt on various aspects of the normative ethics at issue in this article. Among other things, some deny that there can be an overall goodness of a state of affairs; others

(1996), pp. 161–177; Jesper Ryberg, 'The Repugnant Conclusion and Worthwhile Living', *The Repugnant Conclusion: Essays on Population Ethics*, ed. J. Ryberg and T. Tännsjö (Dordrecht, 2004), pp. 239–56; Torbjorn Tännsjö, 'Who are the Beneficiaries?', *Bioethics* 6 (1992), pp. 288–96; Torbjorn Tännsjö, 'Why We Ought to Accept the Repugnant Conclusion', *Utilitas* 14 (2002), pp. 339–59, reprinted in *The Repugnant Conclusion*, ed. Ryberg and Tännsjö, pp. 219–38.

[3] See Thomas Hurka, 'Value and Population Size', *Ethics* 93 (1983), pp. 496–507.

[4] See Larry S. Temkin, 'Intransitivity and the Mere Addition Paradox', *Philosophy and Public Affairs* 16 (1987), pp. 138–87.

[5] There are different views about what precisely counts as 'a utilitarian approach to population ethics' or 'a utilitarian account', and there is not an obvious fact of the matter. Here and in what follows, I leave this definitional question largely untouched, and, when I refer to accounts as 'utilitarian' or 'non-utilitarian", I mean to refer to paradigmatic cases of each – rather than cases at the margins where an account's utilitarian or non-utilitarian status is controversial. For one interesting discussion on this topic see Parfit, *Reasons and Persons*, app. I.

https://doi.org/10.1017/S0953820815000072 Published online by Cambridge University Press

doubt the authority of intuitions in providing ethical knowledge.[6] In this article, however, I will bracket these issues and assume that there is such a thing as overall goodness and that our intuitions can provide ethical knowledge. Next, I will be assuming a utilitarian account for the assessment of the goodness of states of affairs. This does not mean that the discussion will not be of interest to a non-utilitarian. Many questions in population ethics, while often couched in utilitarian terms, are equally important and relevant to the production of complete descriptions of other non-utilitarian accounts.

Further, while I will not be addressing issues such as distributional equity, those who consider welfare a component, but not the sole component, of the good can think of the issues discussed here as merely analysis of the welfare component of good – as opposed to the entire picture. Thus, the discussion here is compatible with there being further inquiry into other issues once the inquiry into welfare is complete.

Next, there are two different ways of understanding utilitarianism. One might think of it as a way of making comparisons between the goodness of states of affairs, but another equally common way of thinking of it is as a way of making prescriptions about actions. These notions might not be identical, but surely there is a close connection. The connection I will assume in this article is that the better action, according to utilitarianism, is the action that brings about the better state of affairs. Thus, the better of two actions, according to TU, is the one that brings about the state of affairs with the greater total amount of happiness, and the better of two actions, according to AU, is the one that brings about the state of affairs with the greater average amount of happiness per person. These clarifications notwithstanding, however, my primary purpose in this article – though not my sole purpose – will be to address axiological questions.

Lastly, despite the initial plausibility of what Parfit calls 'person-affecting moralities", the vast majority of the population ethics literature rejects even utilitarian person-affecting theories and instead focuses on 'impersonal' utilitarian theories. According to Parfit, a 'person-affecting' morality is a moral theory according to which the part of morality concerned with human well-being 'should be explained entirely in terms of what would be good or bad for those people whom our acts affect'.[7] As Parfit and countless others argue, however, person-affecting moralities must be rejected because they are unable to solve the non-identity problem.[8] The non-identity problem is illustrated by

[6] John Mackie, *Ethics: Inventing Right and Wrong* (New York, 1977).
[7] Parfit, *Reasons and Persons*, p. 394.
[8] Parfit, *Reasons and Persons*, p. 359.

https://doi.org/10.1017/S0953820815000072 Published online by Cambridge University Press

Parfit's '14-Year-Old Girl' case. A young girl has a child, who, due to her mother's youth, has a life that is difficult, yet worth living. If, however, the girl had waited a number of years to have a child, she would have had a different child, but one that would have had a better life than that of her actual child. The problem is that it seems clear that it would be better if the girl waited before having a child, but we cannot explain this by saying the girl's decision was worse for her child, because if the girl had waited, a different child would have been born.[9] Thus, despite its initial plausibility and strong attempts to defend it,[10] a person-affecting morality seems ill-equipped to handle questions of population ethics.[11] Instead, and for good reason, Parfit and others turn to impersonal moralities – moralities that are not limited to a consideration of what is good or bad for those affected by an act – with the hopes of constructing a plausible account. Both AU and TU are impersonal moralities, and it is to these views that I now turn.

II. THE MAIN INTUITION MOTIVATING AVERAGE UTILITARIANISM

The main intuition motivating AU – and distinguishing it from TU – is the idea that while we want people to be as happy as possible, we do not necessarily have reason to bring into existence additional happy people. In other words, when it comes to happy people, more is not better. This is what Jan Narveson has in mind when he writes: '[T]here is no moral argument at issue here. How large a population you like is purely a matter of taste, except in cases where a larger population would, due to indirect effects, be happier than the first.'[12] Further, Narveson explains that it is not good that people exist 'because their lives contain happiness'. Instead, 'happiness is good because it is good for people'.[13] Similarly, according to David Boonin-Vail, we should aim 'to produce more happiness for people, not to produce more people for

[9] Some would deny that this case gives reason to reject person-affecting moralities, arguing that the second child would be worse off in the state of affairs where the girl does not wait to have a child. However, it seems odd to say that this child is harmed by not being brought into existence, and maintaining this position would require one to accept additional implausible positions as well. For further discussion, see John Broome, *Weighing Lives* (Oxford, 2004).

[10] See Melinda A. Roberts, 'Person-Based Consequentialism and the Procreation Obligation', *The Repugnant Conclusion*, ed. Ryberg and Tännsjö, pp. 99–128.

[11] See Nils Holtug, 'Person-Affecting Moralities', *The Repugnant Conclusion* ed. Ryberg and Tannsjo, pp. 129–62, for a discussion of the difficulty – in fact, the impossibility – of constructing a plausible person-affecting morality.

[12] Jan Narveson, 'Utilitarianism and New Generations', *Mind* 76 (1967), pp. 62–72, at p. 68.

[13] Parfit, *Reasons and Persons*, p. 394.

https://doi.org/10.1017/S0953820815000072 Published online by Cambridge University Press

happiness'.[14] This is straightforward enough, and it might strike many as plausible. This intuition, however, is consistent with at least two different views, which I will now discuss: first, the view that endorses what John Broome calls the 'neutrality intuition', and second, AU.[15]

II.A. The neutrality intuition

The thought that a larger population of happy people is not necessarily better than a smaller population of happy people has led some authors, including Narveson, to adopt what Broome calls the 'neutrality intuition'. Narveson appeals to this intuition when he further clarifies his position described above: 'We are in favor of making people happy, but *neutral* about making happy people.'[16] In other words, not only does Narveson think that the size of population is a non-moral matter, but he thinks that adding an additional person to a population is something that we should generally consider to be an ethically neutral event. It is neither good nor bad. Narveson himself, however, holds an asymmetrical neutrality intuition. He thinks that the neutrality does not apply to individuals whose lives are comprised of net suffering – he thinks the addition of such a person would not be neutral, but a bad thing. Broome further qualifies what he takes to be the commonly held neutrality intuition. He suggests that there also might be a sufficiently high level of happiness in a potential person that would result in the creation of this person being not neutral, but a good thing. Nevertheless, the thought is that there exists at least a range of happiness levels, which, if possessed by an additional person, would render the creation of that person a morally neutral event.

As Broome points out, however, although the neutrality intuition might be appealing, it can be shown to imply a contradiction, suggesting that it is in fact incoherent. Consider the three following possible worlds, in Example 1. (Each number represents the happiness level[17] of an individual in the world. The 'x' represents the fact that there is no individual who exists in this slot.)

[14] David Boonin-Vail, 'Don't Stop Thinking About Tomorrow: Two Paradoxes about Duties to Future Generations', *Philosophy and Public Affairs* 25 (1996), pp. 267–307, at p. 268.

[15] Broome, *Weighing Lives*.

[16] Jan Narveson, 'Moral Problems of Population', *The Monist* 57 (1973), pp. 62–86, at p. 73.

[17] Throughout the article, reference will be made to 'happiness levels' and 'quantities of value'. While there are many interesting and important issues surrounding how to determine what happiness level an individual has and what these numbers actually represent, this article will attempt to bracket these issues. I will simply make the seemingly plausible assumption that sense can be made of numerical values representing people's happiness.

https://doi.org/10.1017/S0953820815000072 Published online by Cambridge University Press

Example 1[18]

$$A = (1, 1, x)$$
$$B = (1, 1, 1)$$
$$C = (1, 1, 2)$$

According to the neutrality intuition, A is equally as good as B, and A is equally as good as C, yet C is better than B (since it is agreed that it is a good thing, all else equal, to maximize the happiness of members existing in a population). This appears to be a devastating implication of the neutrality intuition. One might, however, try to avoid this implication by espousing a slightly different version of the neutrality intuition – one grounded in the notion of conditional good.

According to a notion of conditional good, if X is good conditional on Y, this means that (Y and X) is better than (Y and ~X), but that neither (Y and X) nor (Y and ~X) is better than, worse than, *or equally as good as* either (~Y and X) or (~Y and ~X). In the context of our topic, the idea is that the goodness of the third person's happiness is conditional on his existence. This avoids the contradiction discussed above, because this account denies that A is equally as good as either B or C – in addition to claiming that A is neither better nor worse than either B or C.

Unfortunately, however, despite avoiding the contradiction, conditional good does not help us much. It fails to give us a way to make comparisons between worlds with different numbers of people. While this might not be deemed particularly problematic in the case of Example 1 – in fact, it illustrates the commonly held neutrality intuition – it seems more problematic if the individuals who exist in both worlds are not equally well off. In other words, it's not clear what the conditional good account would say about a comparison between

$$A = (1, 1, x) \text{ and } D = (1, 2, 1).$$

Further still, the inability to make comparisons between worlds with different numbers of people is particularly problematic if we attempt to compare worlds that do not have individuals assigned to particular slots – for example, if none of the individuals in one world exists in the other, and vice versa. Consider a comparison between worlds E and F, where E includes a '4' and a '5', and F includes a '4', a '6' and a '1'. Depending on how the neutrality intuition account addresses the comparison between A and D, above, two equally unattractive responses could be given with respect to the comparison between E and F. Either the assessment

[18] This is a very slight variation of an example offered by Broome, *Weighing Lives*, p. 146.

https://doi.org/10.1017/S0953820815000072 Published online by Cambridge University Press

of which world is better would seemingly depend on which value, for calculation purposes, is put in the third slot,[19] or no comparison at all can be made since there are different numbers of individuals in each world. Neither of these replies is attractive, however, and the latter can be shown to be particularly absurd in cases where all individuals in the populous world are drastically better off than those in the less populous world.

There are various other ways of attempting to defend the neutrality intuition, but they all are afflicted by serious difficulties.[20] The underlying problem seems to be that the neutrality intuition is a component of a person-affecting morality, and while there are situations that make it seem like a plausible account of population ethics, it is unable to provide a theory that can satisfactorily address the vast majority of cases.

II.B. Average utilitarianism

Despite the fact that the neutrality intuition appears to be false and that the notion of conditional good fails to vindicate it, this does not mean that we are forced to accept the claim that a greater quantity of happy people, all else equal, is better than a lesser quantity of happy people. In fact, although Narveson's and Boonin-Vail's quotations explicitly and implicitly endorse person-affecting moralities, and although the anti-more-is-better intuition will likely be appealing to proponents of person-affecting moralities, the anti-more-is-better intuition *does not* rely on a person-affecting morality. Even if we reject person-affecting moralities and the two related accounts discussed above, we can still largely maintain the core intuition that happiness is only a good conditional on a person's existence. We can do this by appealing to AU.

[19] It does not seem as though the arbitrary choice of how to order individuals should have an effect on the betterness ranking of two worlds, and in cases such as the comparison between E and F, it seems that it would. If the conditional good account allows for the comparison between E and F, it seemingly would say that (4, 5) is better than (4, 1, 6), but worse than (4, 6, 1). One might think that if, among the arbitrary orderings of individuals within each world, there are orderings that provide different prescriptions, then perhaps this indicates that the two worlds are equally good. However, not only does this inference seem not to cohere with the intuitions that result in the seeming contradiction, but there would be extreme cases that could be constructed that would make this rule seem implausible. For example, consider a comparison between A (2) and B (9, 9, 9, 9, 9, and 1.99). One of the arbitrary orderings of B would put 1.99 in the first slot, and since this means that there would be at least one ordering for a prescription in each direction, the result, on this rule, would then be that A is equally as good as B.

[20] Among the approaches that have been offered are views that (1) deny the transitivity of 'equally as good as", (2) maintain that goodness is relative and not absolute, (3) deny that betterness is fully determinate, and (4) no longer try to find an axiological account of the neutrality intuition. See Broome, *Weighing Lives*, p. 149, for further discussion of the notion of conditional good, and of the failure of these other approaches.

https://doi.org/10.1017/S0953820815000072 Published online by Cambridge University Press

Consider Example 1. If we assess goodness of a state of affairs by the degree to which it maximizes average happiness, we get a similar result in the sense that the non-existence of an additional happy (third) person in World A does not necessarily count against its value. However, using AU, we have guidance about which of the three worlds to choose, because there is no intransitivity or other difficulty that results during the comparison between worlds. Each world can be attributed a fixed value that can be compared with the values of other worlds. This is the case even when it's not determined which particular person is occupying the third slot.

AU is not, however, simply an improved and more coherent version of a position embracing the neutrality intuition. In fact, despite their similarities, AU rejects the neutrality intuition.[21] Consider Example 1. A position embracing the neutrality intuition would say that A is equally as good as (or incomparable with) both B and C. AU, on the other hand, will say that C is better than A because the third person raises the average happiness of C above the average happiness of A. AU will still say that A is equally as good as B because the third person does not raise or lower the average happiness of B in relation to that of A. Nevertheless, AU allows us to maintain the intuition that a greater quantity of happy people, all else equal, is not better than a lesser quantity of happy people.

It is important to note, at this point, that while I think that many of those who reject AU will hold the anti-more-is-better intuition but then abandon AU because of one or more of the three objections that will be discussed in this article, there are also people who *do not* hold the anti-more-is better intuition. Thus, to persuade someone in this camp of the truth of AU (over, perhaps, TU), one would need to do more than show that AU is the theory best able to maintain the anti-more-is-better intuition and that the three objections to AU can be met. Additional support for AU would be needed, because, at best, the above might render one agnostic between, perhaps, AU and TU. This support exists, however. While perhaps TU is the first version of utilitarianism that one might think of, it is confronted with serious objections – namely the Repugnant Conclusion – and the same is true for other theories. Thus, this alone gives us reason to look for a different utilitarian theory, and this is enough reason for a non-holder of the anti-more-is-better intuition to take interest in this article's project.[22]

[21] In fact, the rejection of the neutrality intuition is what enables AU to be coherent in the ways discussed in the previous paragraph – ways in which a position embracing the neutrality intuition would not be coherent.

[22] One might think that the 'anti-more-is-better intuition' – the foundation of my argument for AU – begs the question in a debate with a proponent of TU. In a sense

https://doi.org/10.1017/S0953820815000072 Published online by Cambridge University Press

One final point is in order before continuing. Average utilitarianism is a term that does not refer to one view, but, rather, it refers to a family of related views, and at least three of these are worth briefly mentioning.[23] AU1 says that the value of a life is the sum (or integral) of momentary happiness over time, and that the value of a world is the average value of the lives in that world. AU2, on the other hand says that the value of a world is the average value of the person-time-slices that exist. AU3 says that the value of a life is the average value of the moments in that life, and the value of a world is the average value of the lives in it. This is the same as AU2 if every life has the same length, but otherwise it gives greater weight to moments in shorter lives. While AU2 and AU3 are interesting views, they are in conflict with the strong intuition that, all else equal, a longer happy life is better than a shorter happy life, and – although I will not argue that AU1 is the most plausible of the three – I will not further consider AU2 or AU3 in this article. As AU1 is the version people generally refer to in discussions of AU, it is likely to be the one that will strike readers as least implausible, and it is what I will simply call AU for the remainder of this article.[24]

III. OBJECTIONS TO AVERAGE UTILITARIANISM

Why is it, then, that a seemingly coherent view that seemingly plausibly rejects the 'creating more happy people, all else equal, is good' view is rejected by most philosophers? While various objections to AU are made, despite their differences, they all seem to be somewhat related to a single concern: assuming that a particular action (e.g. bringing into existence a person with happiness level five) has no effects on others, why should the goodness of the action depend on anything extrinsic to those affected by the action (e.g. the happiness of the rest of society)? AU

this is true, but in a harmless sense. As I say in the text, I think that the majority of proponents of TU are prima facie sympathetic to the anti-more-is-better intuition, but that they then abandon AU because they are persuaded by one or more of the main objections to AU that I will discuss. Thus, in this sense, I am not begging the question, because the target of my arguments is someone who would buy my premise – and not a committed TU proponent. As I mention in this paragraph, though, my argument does not rest on the anti-more-is-better intuition, and it could entice a reader to believe AU even if they do not share this intuition prima facie. This is because there are prima facie difficulties with other views – such as the Repugnant Conclusion.

[23] Each of these could, in turn, be further precisified by stipulating which temporal population is relevant.

[24] Interestingly, it is worth flagging that some object to AU1, arguing that it is susceptible to an intrapersonal analogue of the Repugnant Conclusion. This point could suggest one of two things. First, it could suggest that we should espouse a different version of AU. Second, it could cast doubt on AU1's rejection of TU. This inquiry is beyond the scope of this article, but I do believe that there are plausible replies to both of these concerns.

https://doi.org/10.1017/S0953820815000072 Published online by Cambridge University Press

seems to dictate this conclusion, because the creation of an individual at level five would seemingly be bad if the prior population's average happiness were level ten, but it would be good if the prior population's average happiness were level three.

This concern is manifested in two different types of objections to AU – one that I'll call 'bottom-up' and the other that I'll call 'top-down'. First, the bottom-up approach. One might point out that it is not at all clear which pool of individuals one should strive to maximize the happiness of: all sentient beings who ever existed? All people who ever existed? All people who currently exist? All people who exist currently or will exist in the future? It is unclear whether a non-arbitrary selection of any of these can be made. Further, and more importantly, according to this characterization of the concern, no choice of a relevant temporal population will yield plausible results.

The top-down approach to objecting to AU is to say that AU fails because it does not always maximize what is of intrinsic value. Since, according to utilitarianism, happiness is of intrinsic value and AU seeks to maximize something which often does not maximize the amount of happiness, AU fails to correspond with our intuitions. Michael Huemer provides a top-down argument of this sort.

I will first describe the bottom-up approach, and I will then proceed to describe Huemer's argument, which exemplifies the top-down approach.

III.A. Parfit and the bottom-up approach

Parfit argues that we should reject AU, and he uses the bottom-up approach to explain why.[25] He considers four temporal populations that we might consider to be the relevant ones, and he concludes that AU yields implausible results for each of them. While Parfit also offers his Two Hells objection as a reason to reject AU, I will defer a discussion of Parfit's Two Hells objection until section VI, because it raises a different concern for AU.

Parfit first considers the 'temporally neutral' version of AU, according to which it is best to maximize the average happiness of all people who ever existed. Parfit thinks that this is absurd, because, as Jeff McMahan's 'Egyptology' example shows, a temporally neutral version of AU implies that whether one should have a child depends on facts about all previous lives, including those of the ancient Egyptians. He writes: 'If the Ancient Egyptians had a very high quality of life, it is more likely to be bad to have a child now.'[26] But, as Parfit says,

[25] Parfit, *Reasons and Persons*, p. 420.

[26] Parfit, *Reasons and Persons*, p. 420, citing Jeff McMahan, 'Problems of Population Theory', *Ethics* 92 (1981), pp. 96–127.

https://doi.org/10.1017/S0953820815000072 Published online by Cambridge University Press

'research in Egyptology cannot be relevant to our decision whether to have children'.[27]

Next, Parfit considers the possibility that what AU requires is that we maximize the average happiness of those who live after the action in question. This, however, dictates that, all else equal, we would bring about a better outcome by killing everyone except for the happiest people who are currently living. This, too, seems absurd.

Third, Parfit considers the possibility that what AU requires is that we maximize the average happiness of all those who currently exist and all those who will exist in the future. This, however, according to Parfit, leads to what I will call Reverse Egyptology – because of its relation to McMahan's Egyptology concern. The concern here is that according to a current/forward-looking version of AU, whether it is good or bad to have an additional child who will experience a particular level of predicted happiness will depend on the predicted levels of happiness of individuals who will live in the distant future. The happiness of these individuals in the future, however, strikes Parfit as irrelevant to the goodness of a family's current procreation decision. The decision can be illustrated as follows:

Possible Future 1[28]			*Possible Future 2*		
Have the child:	5, 5, 6, 6 | 18	AU = 8	Have the child:	5, 5, 6, 6 | 1	**AU = 4.6**
Don't have the child:	5, 5, 5, x | 1	**AU = 8.25**	Don't have the child:	5, 5, 5, x | 1	AU = 4

Thus, according to a current/forward-looking AU, it is better not to have the child if Future 1 will occur, but it is better to have the child if Future 2 will occur.

Lastly, Parfit and Huemer both consider a version of AU that includes only currently existing individuals. According to them, even AU including just this temporal population gets the wrong result in Parfit's case of 'Mere Addition'[29] and Huemer's similar case of 'Benign

[27] Parfit, *Reasons and Persons*, p. 420.

[28] This example does not depict the individuals 'who will live in the distant future' as being very far into the future – it does not depict a string of values coming after the fourth value and before the fifth. This, however, is just for the sake of simplicity of presentation, and the point is the same. Also, note that the third individual is slightly better off in the cases where the child is had. This is supposed to represent the happiness a parent gains by having the child. A similar point, however, would apply even if the parent was not depicted as being made better off by having the child.

[29] Parfit, *Reasons and Persons*, p. 419.

https://doi.org/10.1017/S0953820815000072 Published online by Cambridge University Press

Addition'.[30] These two cases, versions of which are depicted below, involve a comparison between world A, where one person exists, and world B, where the same person exists in addition to other individuals who are less well off, but who nevertheless have lives worth living. 'Addition' is not meant to signify a temporal process, but just the imagination of a different world, similar to the first, but which includes additional people. According to AU, A would be better than B, but Parfit and Huemer think that AU gets the wrong result.

Mere Addition	*Benign Addition*
A: (8, x, x, x)	A: (8, x, x, x)
B: (8, 5, 5, 5)	B: (9, 5, 5, 5)

Thus, according to Parfit and countless others, any specification of AU will fail.

III.B. Huemer and the top-down approach

As discussed above, the top-down approach characterizes the problem with AU as follows: while happiness is intrinsically valuable, AU seeks to maximize something else that is not intrinsically valuable, and often at the expense of maximizing total happiness, which is intrinsically valuable. Huemer raises this concern with AU, although he frames it slightly differently.[31] He seeks to show that if happiness is what is of intrinsic value, then it follows that average happiness is irrelevant, and maximizing average happiness does not necessarily lead to maximizing total value. Thus, Huemer takes a step-by-step approach to show why TU follows if happiness is of intrinsic value.

In his article defending TU against threats from the Repugnant Conclusion, Huemer offers his 'Equivalence Argument', which aims to show that if one believes that it is good, all else equal, to maximize the sum of happiness in a single life, then one is committed to believing that we should maximize total happiness in a population, regardless of how low the average happiness in the population is. Huemer's argument hinges on the claim that a world, G, in which a single individual, Sue, comes into existence for ten minutes at a happiness level of ten is equally as good as a world, H, in which two separate individuals, Sue and Mary, come into and out of existence, one after the other, for five minutes each at a happiness level of ten. The happiness in question is the result of eating a sandwich.

[30] Huemer, 'In Defence of Repugnance'.
[31] Huemer, 'In Defence of Repugnance'.

https://doi.org/10.1017/S0953820815000072 Published online by Cambridge University Press

According to Huemer, 'The intrinsically good-making feature of the sandwich-eating experience – namely its pleasureableness – is equally present in either case.'[32] He continues:

> It seems that the two temporal halves of world G are equally good, since each consists in Sue's enjoying an equally good benefit for five minutes. It is hard to think of a reason why one of these five-minute events should be better than the other. Similarly, it seems that the two temporal halves of world H are equally good. Sue and Mary are qualitatively indistinguishable people who enjoy equal benefits for equal lengths of time. So it is hard to see how one of their lives would be better than the other. Finally, it seems that the first half of world G and the first half of world H are equally good. Both consist in Sue's enjoying the same benefit for the same length of time. If the intrinsic value of an event supervenes on its intrinsic, non-evaluative properties, then the first halves of G and H must have equal intrinsic value.[33]

Thus, Huemer reasons that VG2 = VG1 = VH1 = VH2,[34] and thus that worlds G and H are equally good. Therefore, he says that the duration of welfare for an individual is 'evaluatively equivalent to widespreadness of welfare – the number of people enjoying the benefit'.[35] Thus, Huemer thinks that if happiness is what is of intrinsic value, then TU follows.

This is the juncture at which others have abandoned AU and chosen to pursue either TU or variations of TU, focusing their efforts on how TU should handle the Repugnant Conclusion. I offer a different approach. I think that there is something right about AU, and I think it can be preserved as a plausible account of population ethics. Below, in section IV, I reply to Huemer's top-down argument, and then, in section V, I address the bottom-up objections that Parfit and others raise.

IV. REPLYING TO THE TOP-DOWN CONCERN: A DIFFERENT ACCOUNT OF INTRINSIC VALUE

IV.A. The reason Huemer's argument fails

The alleged problem with AU as it stands – the divergence between what it seeks to maximize and what is valuable to us – is due not to a failure of AU, but rather to our mistaken assessment of what has intrinsic value. Just as a new account of what has intrinsic value can

[32] Huemer, 'In Defence of Repugnance', p. 920.

[33] Huemer, 'In Defence of Repugnance', p. 921.

[34] 'V' means value, the second letter denotes the world, and the numeral denotes the temporal part.

[35] Huemer, 'In Defence of Repugnance', p. 921. In addition to arguing that duration of welfare is equivalent to widespreadness of welfare, Huemer also argues that intensity of welfare is equivalent to duration of welfare (and thus to widespreadness of welfare). For the purposes of our discussion here, however, we need not discuss this other argument.

https://doi.org/10.1017/S0953820815000072 Published online by Cambridge University Press

eliminate this divergence, it also prevents Huemer's argument from getting off the ground.

Huemer explicitly says that his argument for TU only follows 'if the intrinsic value of an event supervenes on its intrinsic, non-evaluative properties'.[36] So, does the intrinsic value of an event supervene on its intrinsic, non-evaluative properties? It seems as though the answer is 'yes' if happiness is itself intrinsically good. This has been the assumption, and I agree that Huemer is right if this is the case. On my account, however, a moment of happiness is not itself intrinsically valuable, and as such, Huemer's crucial steps – that VG1 is equal to VH1, and thus that VG is equal to VH – do not follow. Thus, if happiness is not what is of intrinsic value, TU does not necessarily follow.

In short, then, although Huemer does not technically assume his conclusion, he comes close to doing so. He is not assuming his conclusion, because his conclusion differs from his assumption, and it requires some steps to arrive at the conclusion. This notwithstanding, I say it comes close to assuming its conclusion, because an AU theorist will likely think that Huemer's assumptions get him almost all the way to his conclusion, and thus they will reject his assumption that happiness is of intrinsic value – and also his broader assumption that the intrinsic value of an event supervenes on its intrinsic non-evaluative properties.

But, in light of the fact that it *does* seem to be commonly thought that happiness is intrinsically valuable, what reason does the AU theorist have for rejecting these two assumptions? It is to an AU theorist's account of intrinsic value that I now turn.

IV.B. AU's account of intrinsic value

An AU theorist should hold that nothing has *unconditional* intrinsic value, but that a person's happiness has intrinsic value conditional on his existence. In order to assess these claims, let's consider first how they apply to an existing person, and next, how they apply to a potential person.

In terms of intrapersonal motivation, this account prescribes the same self-interested behaviour as would an account on which happiness had unconditional intrinsic value. Both would prescribe that an existing individual maximize his happiness. Are both accounts equally plausible for the intrapersonal context? As Mill argues, the best proof of something's desirability is the fact that we tend to desire it.[37] Given that the individual in question does exist, though, the 'conditional' clause is satisfied and what remains is that, for the individual, happiness

[36] Huemer, 'In Defence of Repugnance', p. 921.
[37] John Stuart Mill, *Utilitarianism* (Indianapolis, 1979/1861).

https://doi.org/10.1017/S0953820815000072 Published online by Cambridge University Press

is intrinsically good. Thus, the conditional value and the unconditional value accounts are each plausible in an intrapersonal context, and they each pass Mill's test equally well.

As for the case of creating additional people, the two accounts of intrinsic value come apart. While the account according to which happiness is an unconditional good would espouse bringing about additional individuals if they are happy, the account according to which the value of happiness is conditional on existence will not espouse bringing about additional individuals, maintaining that population size is not a moral matter, but rather a matter of taste.

Of course, a proponent of TU will find the unconditional account plausible because he will think it gets the right results in both of the above cases. An AU proponent, on the other hand, will think that the conditional account is plausible because he will think that it gets the right result in both cases. In light of this, we reach an impasse, and it seems that the debate about what, if anything, constitutes an intrinsic good remains split along AU/TU lines. As such, this is why Huemer's argument was described as coming close to assuming its conclusion, and this is why it should not shake anyone from being a proponent of AU. The AU proponent's account of intrinsic value lines up with AU's claims, and there is no need for him to accept the TU proponent's account of intrinsic value.

If a person's happiness has intrinsic value conditional on his existence, a question then becomes what precisely the AU theorist would say about the case in Huemer's equivalence argument. The AU theorist will say that world G (the world with just one person alive for ten minutes) is better than world H (where two people are alive for five minutes each), because G2 is better than H2. Despite the fact that G2's and H2's intrinsic, non-evaluative properties are the same, they differ in that G2 is experienced by an individual (Sue) who has already been alive and experienced happiness, whereas H2 is the first and only experience that Mary has. Thus, if the five minutes of happiness were added to Sue's life, the existence condition would have been satisfied, and the units of happiness would thus be of intrinsic value. However, creating Mary to experience the five minutes of happiness would not add intrinsic value to the state of affairs, because Mary did not exist before this decision. Thus, for an AU theorist, the non-intrinsic properties of the experience of G2 or H2 are relevant to the assessment of the goodness of a state of affairs.

Further, not only does AU's manner of aggregation lead one to adopt the above account of intrinsic value, but the above account of intrinsic value leads one to adopt AU's manner of aggregation. In light of the above claims about the relevance of non-intrinsic properties of

https://doi.org/10.1017/S0953820815000072 Published online by Cambridge University Press

experience, the natural way to aggregate value will be to perform an averaging function over each individual in the population. There are different ways in which these values could be averaged, but I think the most straightforward and plausible one is the one that weights equally the lives of each individual. Regardless of the fact that there are other possibilities here, the various methods of averaging would all come out in favour of AU and against TU. The value of a population, on this account, is its average amount of happiness experienced per life.

IV.C. Summary

The crux of the top-down characterization of the concern with AU before I introduced this account was that maximizing average happiness does not track the maximizing of what is intrinsically valuable. If the current account of intrinsic value is correct, though, this precise problem has now been eliminated, because the AU theorist does not espouse the account of intrinsic value that Huemer assumed that the AU theorist would.

Importantly, however, the reader might recall that, as discussed in section II.A, the notion of conditional good does not provide us with the tools necessary to attack problems in population ethics. This does not mean, however, that it is unable to contribute to the foundations of an account that *could* be successful if armed with additional tools. That is precisely what is occurring here: conditional good helps explain average utilitarianism's account of intrinsic value, but average utilitarianism is able to employ further tools as well. There is nothing inconsistent about this.

There is, however, a concern that one might have at this point: I have not yet articulated an answer to the question of which population is the relevant population for which to maximize average happiness, and the attempt to do so might reveal that AU – and thus, AU's account of intrinsic value – is implausible. It is this question to which I now turn.

V. REPLYING TO THE BOTTOM-UP CONCERN: IDENTIFYING THE RELEVANT POPULATION

One of the main difficulties with AU that Parfit addresses, as discussed in section III, is the fact that it seems to be difficult to come up with a plausible account of the relevant temporal population for which to maximize average happiness. Ultimately, all of the different populations considered seemingly fail and this is what leads to the conclusion that AU fails as a theory – forcing many philosophers to abandon the intuition that more happy people, all else equal, is not better.

https://doi.org/10.1017/S0953820815000072 Published online by Cambridge University Press

In this section, I will begin by redefining the challenge for AU as a need to articulate not a plausible temporal account, but rather, a plausible account of whether or not unaffected individuals should be included in the relevant population. Next, I explain that the question of which population is relevant is distinct from the question of how to assess what is best for that population – as such, TU is confronted with the question as well. I suggest that an impersonal morality should both be temporally neutral and include unaffected individuals. I then take up Benign Addition, Reverse Egyptology, and Egyptology, and I argue that the fact that people find these objections persuasive is not evidence that TU is more plausible than AU, because their persuasive power is due to our person-affecting intuitions and the fact that person-affecting intuitions align with TU in these cases. I argue that there are just as many cases where person-affecting intuitions align with AU and thus, if we are to compare AU and TU, we should do so impersonally. Once we do so, the anti-more-is-better intuition comes into play, without much opposition, and it is difficult to reject AU.

V.A. Redefining the challenge for AU

While Parfit and seemingly everyone else who discusses the question of the relevant population for AU frame their discussions in terms of temporal populations, this is not the best way to address the issue. Rather, the better question is which temporal population of individuals *unaffected by the action* in question should be included in the relevant population, in addition to those, in any temporal population, who are either affected by or brought into or out of existence by the action in question. (Hereafter I will refer to the group of those who are either affected by or brought into or out of existence by the action in question as individuals that are 'not-unaffected'.) Everyone seems to think that not-unaffected individuals should be included in the relevant population for AU, regardless of when they exist. Parfit's own discussion of strategies of conservation and depletion,[38] among others, indicates that there does not seem to be anything counterintuitive about including future individuals' happiness in the relevant population if they are not-unaffected.

What the relevant populations addressed by Mere Addition, Reverse Egyptology and Egyptology have in common is that they include a group of not-unaffected individuals. They differ with respect to the particular temporal population of unaffected individuals that they include: Mere Addition includes unaffected individual in the present, Reverse Egyptology includes unaffected individuals in the present and

[38] Parfit, *Reasons and Persons*, p. 361.

https://doi.org/10.1017/S0953820815000072 Published online by Cambridge University Press

in the future, and Egyptology includes unaffected individuals in the present, future and past. Thus, while Parfit and others are right to suggest that the alleged lack of appeal of AU is due to the inability to articulate a plausible temporal population for which to calculate average happiness, what people actually find unappealing about the various populations considered is that they include individuals who are unaffected by the action in question.

Further, what is allegedly unappealing about including unaffected individuals in the relevant population seems to be equally the case for all of the unaffected individuals, regardless of whether they exist in the past, present or future. Thus, while, above, I stated that a better question than asking which temporal population is relevant is to ask which temporal population of *unaffected* individuals is relevant, the best question seems to be whether the relevant population should include only not-unaffected individuals or whether it should include both not-unaffected individuals and unaffected individuals. If any unaffected individuals should be excluded, there seems to be no reason to include those from other temporal populations, and vice versa. While perhaps the past seems less able to be affected than the future or the present, what is relevant is that it has been stipulated that the future and present individuals in question would be unaffected by the current action as well.

Thus, the challenge still remains for AU, but the challenge seems to be that it must articulate whether the relevant population for which to maximize happiness includes both unaffected individuals and not-unaffected individuals, or whether it only includes not-unaffected individuals.

V.B. Both AU and TU must articulate the relevant population

It is important to note something that seems to have been somewhat ignored[39] in the population ethics literature that is critical of AU: this task of articulating the population relevant for ethical inquiry is not a question only for AU. It is also a question for TU – AU and TU are both impersonal moralities that seek to maximize an abstract value for a population. As such, the question of which population is relevant for ethical inquiry is a question that is distinct and separable from the question of how to assess what is best for that population.

[39] There certainly has been discussion about the relevance of future persons, merely possible persons, and persons that are contingent relative to the set of options, but the fact that TU must answer the question of which population is relevant does not seem to get mentioned in critiques of AU. Discussions of AU's shortcomings do not make clear that the choice of the relevant population is a question that should be answered *prior* to the debate between AU and TU, and thus should be answered without being influenced by a thinker's commitment to AU or TU.

https://doi.org/10.1017/S0953820815000072 Published online by Cambridge University Press

Just as an AU account is incomplete without a description of the relevant population, so too is a TU account. TU's common prescription to 'maximize happiness' or to 'maximize happiness of the population in question' leaves undetermined which population's happiness should be maximized.

Of course it is no surprise that the task of articulating the relevant population is thought to be a task only for AU. The inclusion of various temporal populations of unaffected individuals will often affect AU's prescriptions, but for TU, these unaffected individuals in any temporal population will just serve as a constant that does not affect TU's prescriptions. Thus, while examples like Parfit's case of depletion and conservation show that it matters to TU which not-unaffected individuals are included in a population, it seems as though TU does not confront a question of whether unaffected individuals are relevant. In the cases of Benign Addition, Egyptology and Reverse Egyptology, no matter how many unaffected individuals are included in the relevant population, TU's prescription about whether the individual in question should procreate will remain the same.

Nevertheless, TU is an impersonal morality that seeks to maximize happiness for a population, and as such, it is left under-described if it does not articulate whether unaffected individuals are part of the relevant population for ethical inquiry. Even if different versions of TU provided identical prescriptions, it would be important to articulate which version one subscribed to. This notwithstanding, the fact that TU is confronted with the same task of selecting the relevant population as is AU does not mean that AU is off the hook. If TU were plausible for at least one population while AU were not plausible for any population, this would indicate that TU is a plausible theory and that AU is not. Nevertheless, the question of which population is relevant and the AU versus TU question are conceptually independent.

V.C. The relevant population should include unaffected individuals and unaffected values

Up to this point I have used the term 'unaffected individual' as an umbrella concept to refer to two separate cases that I will now distinguish: cases of what I will more properly call 'unaffected individuals' and cases of what I will call 'unaffected values'. I argue that both unaffected individuals *and* unaffected values must be included in the relevant population, and given that the question of relevant population is distinct from the question of AU versus TU, this result will apply to both theories.

A common thought might be that the relevant population for ethical inquiry should exclude individuals who exist in each of two states of affairs and who will be equally well off in each, i.e. 'unaffected

https://doi.org/10.1017/S0953820815000072 Published online by Cambridge University Press

individuals'. However, the relevant population should not be restricted in this way. As a result of the difficulties exposed by Parfit's two-child case, it was necessary to adopt an impersonal morality that would treat the two following cases alike: (1) a case of two different individuals, one in s1 and at value '3' and the other in s2 and at value '7', and (2) a case where an individual exists in s1 at value '3', and where the *same* individual exists in s2 at value '7'. However, if we are going to ignore identity in cases like this, it seems as though an impersonal morality should also treat the following two cases in the same way: (1) a case of two different individuals, one in s1 and at value '5' and the other in s2 and at value '5' and (2) a case where an individual exists in s1 at value '5', and where the *same* individual exists in s2 at value '5'. Thus, if an impersonal morality includes in the relevant population for ethical inquiry two different individuals with the same value (cases of what I call 'unaffected value'), it should also include in the population one individual who is equally well off in both states of affairs (cases of what I call 'unaffected individual'). Further, as I argue below, it seems as though an impersonal morality *should* include unaffected values in the relevant population, and thus, it *should* include unaffected individuals.

Someone who thinks that unaffected values should not be included in the relevant population would think that the way to compare states of affairs would be to consider only the differences between their strings of values. For example, in comparing A and B, below, we might match up four of the '5's in each state of affairs, thus finding that only the 'x' and one '5' in A and the '6' and the '2' in B are the population relevant for ethical inquiry.

While this is perhaps plausible at first, consider the possibility that

A:	5, 5, 5, 5, 5, x
B:	5, 5, 5, 5, 6, 2

the option of bringing about state of affairs C arises, and the decision becomes whether to bring about A, B or C:

C:	3, 7, 3, 7, 8, 1

C does not share any values with A or B, and thus it is clear that all of the '5's that had not been considered part of the relevant population now must be considered part of the relevant population. Thus, while on TU, the comparison between A and B seemingly could leave out the first

https://doi.org/10.1017/S0953820815000072 Published online by Cambridge University Press

four '5's, this is not because these '5's are not relevant to ethical inquiry, but it's just that in this case they happen not to affect the result.[40]

One might say, however, that unaffected values are relative to all available options. However, even if we can say that unaffected values are relative to all available options, this shows that the notion of an unaffected value is not a fundamental part of an impersonal morality. In order to know which values are unaffected, we have to be told what the consequences of the available actions are, and only from this can we deduce which values are affected and which not. Thus, it would be wrong to say that we have some prior notion of which values are affected. Rather, this is a judgement derived from something more fundamental – the raw data about *all* of the values themselves.

The broader point here is that the causal consequence of a choice about how to act is the entire resulting state of affairs, and it is this that must be compared to the causal consequence of a different choice – also an entire resulting state of affairs. Further, as discussed above, the two possible causal consequences of a choice are the same, regardless of the identities of the individuals and how directly or indirectly the agent brings about the various components of the resulting states of affairs.

Thus, an impersonal morality should be temporally neutral and should include both unaffected individuals and not-unaffected individuals.[41]

V.D. Why Benign Addition, Reverse Egyptology and Egyptology all miss the mark

Now that we understand better what TU and AU entail, we are better equipped to see why Benign Addition, Reverse Egyptology and Egyptology miss the mark. I will argue that the fact that people find these objections persuasive is not evidence that TU is more plausible than AU, because their persuasive power is due to our person-affecting intuitions and the fact that person-affecting intuitions align with TU in these cases. To illustrate this, I show that versions of these examples that are slightly altered to produce alignment of person-affecting intuitions with AU result in AU being likely to seem more attractive. Thus, I suggest that the result is a wash and that TU and

[40] In other words, while the following is not the case for AU, TU satisfies what one might call a monotonicity condition, according to which population X is better than population Y if and only if the combined population X & Z is better than Y & Z. However, this property of TU notwithstanding, as just argued, even unaffected values (e.g. population Z) are *still* part of the population relevant to ethical inquiry.

[41] I'm open to the idea of including other conscious creatures as well, but this would not change the structure of the theory.

https://doi.org/10.1017/S0953820815000072 Published online by Cambridge University Press

AU must be compared in a manner divorced from the complicating person-affecting intuitions.[42]

V.D.1. Benign Addition

Parfit and Huemer both argue that Benign Addition and the related cases give us reason to think that TU is more plausible than AU. How, they ask, can it be worse for there to be additional happy people (Mere Addition), and how can it still be worse to add additional happy people when the original person is made better off (Benign Addition)? To avoid confusion, in what follows I will discuss Benign Addition, but similar points apply to Mere Addition.

Benign Addition	*Benign Addition Variation*
A: (8, x, x, x)	A: (8, x, x, x)
B: (9, 5, 5, 5)	C: (5, 5, 5, 9)

The objection says that in Benign Addition, B is better than A, and the idea is that this is because of TU. My thought is that this has nothing to do with TU. Rather, I think this is probably because of the person-affecting intuition, which here suggests that B is better. One way to test this might be to consider Benign Addition Variation. Would it be better to reduce the welfare of an existing individual to bring into existence three additional happy individuals?

AU and TU are both impersonal moralities that seek to maximize average and total happiness, respectively, for the relevant population, and we know that all of the values must be included in the population. Further, the fact that whether to include unaffected individuals in the relevant population does not alter TU's prescriptions seems to have masked the fact that TU is an impersonal morality and not a person-affecting morality. In the above cases, AU and TU will have the same prescriptions in comparisons between A and B and between A and C – with AU saying A is better in both cases and TU saying it is worse in both cases.

[42] I owe thanks to an anonymous reviewer for making the following point and for stressing its importance. Although I refer throughout this article to 'person-affecting intuitions', it is important to recognize the differences between various types of person-affecting intuitions. Among the different possible formulations are the following two. First, x is better than y, *if* (x is better than y for some actual person and worse for no one). Second, x is better than y, *only if* x is better than y for some actual person. Unless stated otherwise in what follows, I will intend a third formulation – the conjunction of the first and second formulations. It is helpful to notice when making comparisons, though, which conjunct of the conjunction is operative in making a particular claim. In most cases, however, it is the second formulation that is doing the work.

https://doi.org/10.1017/S0953820815000072 Published online by Cambridge University Press

While a proponent of TU will maintain that C is better than A, I think that many people who prefer B to A and who think that this is because they find TU plausible[43] will in fact find that they prefer A to C in Benign Addition Variation. If so, it seems as though the alleged force of the Benign Addition example is not a function of intuitions in favour of TU, but rather a function of our person-affecting intuitions – and the same could be said about the force of Benign Addition Variation if it were used to show that AU were more plausible than TU. In Benign Addition, person-affecting intuitions align with TU, and in Benign Addition Variation they align with AU.

V.D.2. Reverse Egyptology

The same thing at work in Benign Addition seems to be at work in Reverse Egyptology as well: person-affecting intuitions align with TU in Reverse Egyptology, but in similar cases that TU and AU would treat in the same way as Reverse Egyptology, person-affecting intuitions align with AU. Recall the Reverse Egyptology example:

Possible Future 1		
1A: Have the child:	5, 5, 6, 6 | 18	AU = 8, **TU = 40**
1B: Don't have the child:	5, 5, 5, x | 18	**AU = 8.25**, TU = 33
Possible Future 2		
2A: Have the child:	5, 5, 6, 6 | 1	**AU = 4.6**, **TU = 23**
2B: Don't have the child:	5, 5, 5, x | 1	AU = 4, TU = 16

The objection aims to draw out our intuitions that, regardless of which of the two possible futures obtains, the same choice should be made – it is better to have the child. This is presented as an objection to AU and as evidence in favour of TU, because TU gives us this prescription, whereas AU does not. AU says that we should have the child if Future 2 will obtain, but that we should not have the child if Future 1 obtains. However, as with Benign Addition, I think that if we have the intuitions that Reverse Egyptology suggests we do, this is not because of TU, but rather because of person-affecting intuitions, which here suggest that having the child is better in both cases. One way to test this might be to consider the following cases:

[43] It is important to note here (and throughout section V.D when I make claims analogous to this) that I am not saying that a committed proponent of TU would find A preferable to C. This is certainly not the case since TU holds that C is preferable to A. Those about whom I'm speaking are those who find B preferable to A and who then conclude that what must be influencing them and what must be doing the work in an explanation of their intuitions is a sympathy for TU.

https://doi.org/10.1017/S0953820815000072 Published online by Cambridge University Press

Case 3		
3A (rearranged values for 1A, above):	5, 5, 6, 6 | 18	AU = 8, **TU = 40**
3B (rearranged values for 1B, above):	5, 5, 5, 18 | x	**AU = 8.25**, TU = 33
Case 4		
4A (rearranged values for 2A, above):	5, 5, 6, 1 | 6	**AU = 4.6**, **TU = 23**
4B (rearranged values for 2B, above):	5, 5, 5, 1 | x	AU = 4, TU = 16

In these cases, two individuals are deciding whether to create a zygote that will be frozen for one thousand years before developing into a child. As the tables show, since there are corresponding values in 1A and 3A, 1B and 3B, 2A and 4A, and in 2B and 4B, AU and TU will have the same prescriptions in this new example. However, I think that many people who claim to prefer TU would find that in case 3 they prefer 3B and that in case 4 they prefer 4A. While 4A is the option TU espouses, 3B is not. Both 3B and 4A, however, are espoused by AU. Similarly, in these two cases, the person-affecting intuitions suggest that 3B and 4A are better. Thus, as with Benign Addition, it seems as though common intuitions in Reverse Egyptology might be due to the fact that, in that example, TU's prescriptions align with the prescriptions of person-affecting intuitions.[44]

V.D.3. Egyptology

So as not to repeat myself, I will not explicitly apply the above analysis to the Egyptology case and the Egyptology variations, depicted below, but the same points apply.

While perhaps it might seem odd to some to include the past in the population relevant for ethical inquiry, recall that I have already argued that those who are unaffected by an action seem to be similarly situated regardless of when they exist temporally. Further, I also argued that all unaffected values should be included in the population relevant for ethical inquiry for an impersonal morality. Even if one accepts this, however, one might still think that there is a difficulty with these examples, because we seemingly would never be confronted with a choice between 3A and 3B or a choice between 4A and 4B. Both seem to suggest that we can change the past. However, since only the values matter for an impersonal morality, it was not crucial that the '18', 'x', '6' and 'x' in the past in Cases 3 and 4 were put in the past – the

[44] In the versions of Reverse Egyptology and Egyptology discussed here, a parent experiences a gain in happiness as a result of having the child in question. However, even if the version analysed were one in which a parent did not experience a gain in happiness, the analysis would be similar, and the conclusion the same: our intuitions in these cases are not dictated by AU or TU, but by person-affecting intuitions.

https://doi.org/10.1017/S0953820815000072 Published online by Cambridge University Press

Happy Egypt		
1A: Have the child:	18 | 5, 5, 6, 6	AU = 8, **TU = 40**
1B: Don't have the child:	18 | 5, 5, 5, x	**AU = 8.25**, TU = 33
Unhappy Egypt		
2A: Have the child:	1 | 5, 5, 6, 6	**AU = 4.6**, **TU = 23**
2B: Don't have the child:	1 | 5, 5, 5, x	AU = 4, TU = 16
Case 3		
3A (rearranged values for 1A, above):	18 | 5, 5, 6, 6	AU = 8, **TU = 40**
3B (rearranged values for 1B, above):	x | 5, 5, 5, 18	**AU = 8.25**, TU = 33
Case 4		
4A (rearranged values for 2A, above):	6 | 5, 5, 6, 1	**AU = 4.6**, **TU = 23**
4B (rearranged values for 2B, above):	x | 5, 5, 5, 1	AU = 4, TU = 16

examples would have sufficiently illustrated the same point about the Egyptology objection even if they were put after the '|'. Further, even if these values were kept before the '|' (and thus in the past), the example maintains its force, because the point being made here is an axiological one – when is x better than y? This question can be asked even if one is unable to make a choice that will bring about x or y, due to x or y being in the past or for other reasons.

Not only is one able to make betterness assessments of this sort, but one can also wish that things in the past were or were not the case. For example, one can wish that it was the case that the ancient Egyptians lived fulfilling lives. Further, and very interestingly, these types of wishes or hopes are not limited to comparisons between possible past events. Parfit's discussion in part II of *Reasons and Persons* illustrates this nicely. First, Parfit plausibly argues (via his Hospital Example and its variations[45]) that when it comes to oneself at a particular time, one will prefer that it be the case that one's life's remainder contains as little suffering as possible, even if this means that one's life will have had a much larger quantity of past suffering, and thus more total suffering during one's lifetime. Despite this, Parfit argues, when it comes to individuals other than oneself – be they near or distant – one wishes foremost not that it be the case that they have the least amount of suffering remaining, but that their entire life (past and future both included) have as little total suffering as possible. What this shows is that, when it comes to thinking about individuals other than oneself, we often hope that future suffering be greater, even if it means that past suffering was less by a greater amount. This provides strong additional support for the claim that the past is and should be included in the

[45] Parfit, *Reasons and Persons*, p. 165.

https://doi.org/10.1017/S0953820815000072 Published online by Cambridge University Press

population relevant for ethical inquiry. Further, while this discussion has focused on axiology, examples can be provided that indicate that even Egyptology can be addressed through the lens of decision-making.[46]

[46] These points should be sufficient, but there is also an explanation of how these points regarding Egyptology can be understood not only as being axiological points, but also as points that apply to the context of decision-making.

Since it still might be difficult for some to accept that past values are relevant for ethical inquiry, I will provide an example to give further intuitive support for this claim – an example which will also show that sense can be made of Cases 3 and 4, even as they are written. This example will illustrate two points:

First, a reader who is sympathetic to evidential decision theory (EDT) will take the example to show that choices between 3A and 3B and between 4A and 4B are choices that can confront us. According to EDT, even TU, in practice, might at times provide different prescriptions based on whether it takes a temporally neutral or current/forward-looking form, and the examples suggests that a proponent of EDT should espouse a temporally neutral morality. This illustrates the important point that any persuasive power that the Egyptology objection provides over and above the persuasive power of Reverse Egyptology is, at best, limited to impressing causal decision theorists.

The example, however, is not introduced only for the eyes of proponents of EDT. Its implications are broader: Even if temporally neutral and current/forward-looking versions of TU never actually have diverging prescriptions, *if they did*, a temporally neutral version would be more plausible. As such, even in cases where the values of individuals in the past are 'unaffected', they are as relevant and important as the unaffected values in the future. Although one might try to draw a distinction between the relevance and importance of happiness that will be experienced and happiness that was experienced, no such distinction can be made.

Genetic Twin Example. You live in a small city, and there's a possibility that you have a gene that causes two behaviours – neither of which would be expressed by many individuals without the gene: (1) in one's thirties, one will work extraordinarily hard to bring about a 20 per cent gain of happiness for each of one's fellow citizens, and (2) in one's fifties one will be destructive and bring about a 80 per cent loss of happiness for each of one's fellow citizens. You know that, a century ago, your genetic twin lived in a city with five times the population of yours. However, you do not know whether he had the particular bad gene, and thus whether he manifested these two behaviours. You do, however, know that he lived to age ninety. You, on the other hand, have been infected with a virus that will kill you in your forties, before you could become destructive to your fellow citizens. What should you do? Should you bring about the 20 per cent gain to your population before your death?

Since bringing about the 20 per cent gain to your population is strong evidence that your brother brought about a net 60 per cent loss to his population, it seems as though your choice is between two states of affairs:

	Brother's city	The present (your city)
A: Help your citizens:	(. . . 10, 10, 10 | 6, 6, 6, 6, 6 | 10, 10, 10 | 12 . . .)	
B: Don't help your citizens:	(. . . 10, 10, 10 | 10, 10, 10, 10, 10 | 10, 10, 10 | 10 . . .)	

Presented with this situation (and assuming EDT), a current/forward-looking version of TU would provide a different prescription than would a temporally neutral version of

https://doi.org/10.1017/S0953820815000072 Published online by Cambridge University Press

Thus, in sum, once the scope is widened beyond not-unaffected individuals, there do not seem to be grounds to exclude from the relevant population any individuals, past, present or future.[47] For

TU. The current/forward-looking version would prescribe helping one's citizens, but the temporally neutral version would prescribe not helping one's citizens: According to EDT, what's relevant is the probability of what occurred in the past conditional on your doing an act. Conditional on your not helping your citizens it's probable that your brother did not have the gene and that his citizens had values of ten. Conditional on your helping your citizens it's probable that your brother did have the gene and that his citizens had values of six. So according to EDT and a temporally neutral morality, one should not help one's citizens. It seems plausible to me that it would be better not to help one's citizens, and thus to bring about state of affairs B. (Of course, while some find this plausible, others find this claim implausible, and thus view this as a *reductio* against EDT. People's intuitions on this are far from unanimous.)

Wherever one's intuitions lie on this question, this example should provide further intuitive support for the claim that past individuals should be included in the population relevant for ethical inquiry – even to a causal decision theorist who denies that a temporally neutral and current/forward-looking TU would, in practice, ever diverge. This is for the following reason: just as we can hope for future individuals to be happy, so too can we hope that past individuals were happy. Thus, even in cases *unlike* the above example, cases where current assessments place identical strings of values for past people in each of the states of affairs being compared, these values should still be taken into consideration by an impersonal morality – be it AU or TU. Once the scope is widened beyond not-unaffected individuals, there do not seem to be grounds to exclude from the relevant population any individuals, past, present or future. As argued earlier in section V, the fact that unaffected individuals in the past cannot be affected does not distinguish them from individuals in the present or future who, it has been stipulated, are unaffected by an act.

[47] An additional concern that one might raise (and that an anonymous referee did raise) for a temporally neutral AU view is that it implies variants of the Repugnant Conclusion. The referee's words are worth providing in full: 'Suppose that, in the past, there existed 10 billion people with lives barely worth living. Each had a lifetime well-being of 1. We have two choices regarding who will exist in the future. In outcome A another 10 billion people will exist, with only marginally better lives than the people in the past. Their lifetime well-being will be 2. In outcome B, 50 million people will exist, with an extremely high lifetime well-being of 100. This means that the average well-being is 1.5 in A, and 1.49 in B. AU therefore favours A. If a main intuition behind AU is "the idea that while we want people to be as happy as possible, we do not necessarily have reason to bring into existence additional happy people" and that "when it comes to happy people, more is not better", the implication that A is better than B seems troubling. If a smaller and very happy total population is preferable to a larger and less happy one, why shouldn't we have this preference also regarding future populations?' This is a reasonable concern that is important to raise. The temporally neutral AU theorist does, however, have the following response: the average happiness must be considered for the whole state of affairs, and choosing to focus on some subset of the entire state of affairs will always enable one to describe cases such as the above. Consider a variation of the case offered, but where the world is just beginning, and both options A and B also include the creation of the original 10 billion people at well-being level 1. In this case, it seems that the AU intuition is that A is better than B. However, as I have argued in this article, the only difference between choice one and choice two is whether the unaffected (and unaffectable!) people are in the past or in the future. It seems that this difference is immaterial to which state of affairs is better, and thus that the two cases should be treated alike. It has been argued that the temporally neutral version is (perhaps counter to initial intuitions) the more plausible version. Thus, as attractive as the intuition in the reviewer's objection may be, it seems that revising it is the least difficult of various routes to reflective equilibrium.

https://doi.org/10.1017/S0953820815000072 Published online by Cambridge University Press

additional explanation of why inclusion of the past is not as implausible as it might initially seem, see the discussion below, in section V.E, about heuristics.

V.E. Summary

While Benign Addition, Reverse Egyptology and Egyptology are examples of cases where commonly held person-affecting intuitions align with TU, I argue that there are just as many cases where they align with AU. Given that these person-affecting intuitions affect our assessments in both directions, if we want to compare AU and TU, we should describe worlds in a manner divorced from these complicating person-affecting intuitions. We should compare worlds without information of whether an individual in one world will exist in the other, and of which value he would have if he did. This information is not relevant to the comparison between AU and TU – we should only consider the strings of values.

Thus, if we are to come to a conclusion about whether AU or TU is more plausible, I think we need to consider this impersonal case. The consideration of this impersonal case might lead some to favour TU and some to favour AU. But this is where the anti-more-is-better intuition (see section II) comes into play, and in my opinion without much opposition. Seemingly everyone who rejects AU – in the context of positive utility – rejects it because of absurdities it allegedly entails in cases where AU conflicts with the person-affecting intuition. Once AU and TU are compared impersonally, as they should be, AU's treatment of cases of positive utility is difficult to reject.

At this point, a few objections might be made. First, one might think that to the extent TU coincides more with person-affecting intuitions than does AU, this is reason to prefer TU to AU. While this might be true if it were the case that TU coincided more with person-affecting morality than does AU, the sets of cases described above are typical. Cases where person-affecting intuitions coincide with TU can be made to coincide with AU by rearranging the slots, and vice versa. As such, there does not seem to be greater overlap between person-affecting intuitions and TU than there is between person-affecting intuitions and AU.

A second objection one might make is that the real take-away here is not about the relative merits of AU and TU, but rather that a person-affecting morality is superior to both. Relatedly, one might say that we should not be abstracting from situations to exclude seemingly relevant facts from our analysis. However, the difficulties that make a person-affecting morality intractable as a theory of population ethics, discussed in section I, are so great that this is not a good option – despite the fact that we might have some person-affecting intuitions.

The fact that we have some person-affecting intuitions, however, can be explained by the fact that in most situations that confront us, these

https://doi.org/10.1017/S0953820815000072 Published online by Cambridge University Press

intuitions are helpful heuristics that simplify our decision-making. Not only are most decisions we make in life examples of decisions in same-numbers cases, but most decisions involve states of affairs involving the same individuals. In decisions of this sort, a person-affecting morality will coincide with both AU and TU, and importantly, focusing merely on what is good or bad for the individuals affected will provide a shortcut for decision-making. Further, given one's limited knowledge about how one's actions might affect the spatiotemporally distant, what is good according to a person-affecting morality, all else equal, will also be what is good on AU or TU. Not only will the treatment of affected individuals serve as a proxy for the current-and-future population that includes both affected individuals and unaffected individuals, but this current-and-future population will serve as a proxy for a temporally neutral population including both affected individuals and unaffected individuals – and, thus, the treatment of affected individuals will serve as a proxy for this temporally neutral population as well. These points notwithstanding, in questions of population ethics, the person-affecting heuristic does not provide the value that it might in everyday cases.

One final note regarding person-affecting intuitions is in order. As with all heuristics, there are cases where the shortcut gets the wrong result, and it is these cases that have been under the microscope in this article. Even in these cases, though, where person-affecting intuitions seem to diverge in their prescriptions from AU, and thus provide the *wrong* answer to moral problems, an AU theorist – or even a TU theorist, for that matter – should not ignore person-affecting intuitions. This is the case because even if the person-affecting intuitions are mistaken, the fact that we have a 'taste' for person-affecting intuitions needs to be taken into account when carrying out calculations for AU or TU. In other words, it will sometimes be best – according to AU or TU – to act in accordance with a person-affecting principle because it is *instrumentally* valuable. If people hold person-affecting intuitions, it might cause real disvalue to act in contravention of these intuitions, and avoiding this disvalue can, at times, be what's best according to either AU or TU.

VI. TWO HELLS

Perhaps the objection to AU that is regarded as the strongest is one that Parfit raised and termed Two Hells.[48] I have yet to discuss it because it raises concerns of a different type from the alleged counterexamples already discussed: it addresses AU's treatment of suffering. The objection goes as follows. Consider two possible hells.

[48] Parfit, *Reasons and Persons*, p. 393.

https://doi.org/10.1017/S0953820815000072 Published online by Cambridge University Press

In Hell 1, ten people exist and suffer great agony. In Hell 2, there are 10 million people who suffer ever so slightly less agony. AU will say that Hell 2 is better. Parfit and countless others, however, find it obvious that Hell 2 is far and away the worse of the two states of affairs.

I am certainly in the minority, but I think that others have been mistaken, and that AU provides an adequate assessment of the Two Hells case. It seems plausible to me that there is nothing wrong with its symmetrical treatment of cases of positive happiness and negative happiness. Narveson and Boonin-Vail, themselves, however, do not think that their slogans about the addition of happy people apply to the addition of unhappy people. But this asymmetry seems unwarranted. With unhappy people, just as with happy people, 'How large a population you like is purely a matter of taste' and 'there is no moral argument at issue here'.[49] It is not bad that unhappy people exist because their lives contain unhappiness; unhappiness is bad because it is bad for people. We should aim to reduce unhappiness for people, not reduce the number of people for a reduction of unhappiness.

This, however, amounts to little less than a bare assertion of the AU theorists' view regarding the Two Hells case. Although I find the above point sufficient, most readers will not, and thus more needs to be said to show that the Two Hells objection is not decisive against AU. As such, the following discussion aims to provide the beginnings of a disarming (or at least of a weakening) of the Two Hells objection to AU.

VI.A. The limited force of the Two Hells example

Although Larry Temkin is not a utilitarian, and although he makes it clear that he finds average utilitarianism to be particularly implausible,[50] I think that a discussion of his, in *Rethinking the Good*, provides fodder for an AU theorist to respond to the Two Hells objection.[51]

According to Temkin, when comparing two states of affairs in which a burden of a very similar quality is imposed on two different groups, the number of people affected in each state of affairs is relevant – or, as he says, 'additive aggregationist' reasoning is plausible. For example, according to Temkin, in a case like Two Hells Hell 2 is much worse

[49] Narveson, 'Utilitarianism and New Generations', p. 68).

[50] Larry S. Temkin, *Rethinking the Good* (Oxford, 2012), pp. 319–23.

[51] As Temkin says, though, the original source of the sort of example that he discusses is Stuart Rachels. Rachels's original example was described in his unpublished Philosophy, Politics, and Economics thesis, 'A Theory of Beneficence' (Oxford University, 1993), and Rachels further discusses the topic in his 'Counterexamples to the Transitivity of *Better Than*', *Australian Journal of Philosophy* 76 (1998), pp. 71–83. In my discussion, I will continue to refer to Temkin's example and discussion, but it should be remembered throughout that Rachels is the original source.

https://doi.org/10.1017/S0953820815000072 Published online by Cambridge University Press

than Hell 1, and this is because the difference in quality of the burdens is very small, and this enables the number of individuals in Hell 2 to make the badness of Hell 2 outweigh the badness of Hell 1. Whether or not the badness of Hell 2 and Hell 1 – at least with respect to the utility ideal, leaving open the possibility that ideals other than utility are valuable – is directly proportional to the total units of utility is an open question, but at the very least, Temkin argues that it is plausible that the number of affected parties is relevant. Further, he argues, a large enough disparity of people can outweigh the slight disparity in quality of the burden that cuts in the opposite direction.

According to Temkin, however, additive aggregationist reasoning is *not* plausible for comparisons of two states of affairs in which there is a large disparity in the quality of a burden. What Temkin means is that in comparisons of this sort, it might be the case that no number of people affected by the lesser burden will make that state of affairs worse than a state of affairs with a particular number of people affected by a severe burden. To use his example, there is no number – however large – of people afflicted by the mild annoyance of mosquito bites that would make such a state of affairs worse than a different state of affairs in which some small number of people undergoes severe torture.

Temkin ultimately argues that the two above claims, combined with a continuity claim (asserting that a sequence of steps between cases with merely slight quality disparities would ultimately get one from the case of torture to a case of mosquito bites), and the claim that the better-than relation is transitive are inconsistent. He laments the need to reject any of these four claims, but he argues that we must reject the transitivity of the betterness relation. There are various reasons to doubt the plausibility of this position, but they are beyond the scope of this article's analysis. Instead, let's focus on the implications of Temkin's first and second claims for the Two Hells objection – assuming that his claims about people's intuitions are fairly accurate, as seems likely to be the case.

Temkin's discussion suggests that while people will generally side with the TU theorist in Parfit's Two Hells case and in similar cases, people will generally side with the AU theorist in cases of suffering where either Hell 1 (or Hell 2) is compared to a Hell 3, where there are an enormous number of individuals each of whom has a utility level that is only slightly negative (e.g. due to mosquito bites or mild headaches). Thus, it seems that comparisons between states of affairs involving suffering do not in fact only support judgements in favour of TU and against AU, but, rather, there will be cases that support the opposite judgement. Further, since it's not clear exactly where people will implicitly draw the line between severe and slight quality

https://doi.org/10.1017/S0953820815000072 Published online by Cambridge University Press

disparities, it's far from clear whether these cases, collectively, would provide more support for AU or for TU. Thus, the point here bears some structural resemblance to the points about reference class made in section V. While it might be the case that intuitions side with TU in the objections given, this is not due to the merits of TU, but rather to the fact that these are cases where the principle that is doing the work aligns with TU, and there seem to be equally many cases where this principle aligns with AU.

As was the case in section V, though, the question then becomes whether this other principle that is more tailored to our intuitions is in fact the principle that should be adopted – instead of either AU or TU. In section V, it was argued that person-affecting principles cannot be adopted because of the great difficulties they face. Similarly here, it seems that we cannot maintain Temkin's two principles about different types of comparisons. Unless one is willing to accept his claims about transitivity, or unless one rejects some other seemingly innocuous claims, it is inconsistent to hold both of his claims about numbers mattering and not mattering in their respective contexts. Thus, it seems that one is in fact required to have a single theory of whether the number of people matters or does not matter. This does not necessarily mean, however, that one must either accept a pure version of TU or AU for the purposes of aggregation. One could adopt a hybrid view of some sort where, perhaps, value is equal to the quality of the burden times the quantity (i.e. the quality times (the quality times the number of people)). This, however, would simply shift the relative weights of quality and quantity, and it would not avoid the problem of Temkin's spectrum.

Thus, ultimately, it seems that one must either reject the absolute claim about the comparisons between cases of large disparity in quality or the absolute claim about the comparisons between cases of slight disparity in quality, or both, and it seems that people will find rejecting either of them to be difficult. As such, it seems that cases of suffering do not provide strong support in favour of TU or in favour of AU. As with cases of positive utility, there are cases where our intuitions support one and there are cases where our intuitions support the other.

VI.B. The relevance of person-affecting intuitions

Given that the context of suffering was supposed to provide strong reason to reject AU alone, it is no small feat to have shown that this context provides potentially equally strong considerations in favour of AU and TU respectively. If the above considerations bring us anywhere close to equipoise between AU and TU, this is a success for the purposes of this article, because this article has proceeded by first motivating

https://doi.org/10.1017/S0953820815000072 Published online by Cambridge University Press

AU, and then attempting to show that the main objections against AU are *not as strong as they otherwise might have seemed*, and not strong enough to merit abandonment of AU. Thus, the discussion of suffering could stop here. Nevertheless, it's worth continuing for two reasons. First, one might think that the intuition in the Two Hells case is stronger and harder to abandon than the intuitions in cases of comparisons involving large quality disparities. Second, even if we truly are in equipoise for the cases of suffering, it's worth hazarding a positive argument (in favour of AU's account of suffering) to attempt to explain people's intuitions in the Two Hells case.

Thus, let's consider a case that is similar to Two Hells, but in which the larger population is at the exact same (negative) utility level as is the smaller population. (It should be noted that this is the type of case that is at one end of Temkin's spectrum, because there is not even a slight quality disparity that favours the larger population.) Further, let's compare this case to its analogue where everything is the same except that the utility values are positive. Though there certainly are readers who will think that the state of affairs with the larger happy population is better than the one with the smaller population, there will be many readers who will be willing to concede that population size is a matter of taste in the positive utility context, but who will deny the analogous claim in the negative utility context. What might explain this?

I think that this difference can be explained by person-affecting intuitions. Let me explain. In the positive utility context, it might be thought that if the additional individuals are brought into existence they are thereby benefited, but that if they are not brought into existence, there is no individual that is harmed. These individuals brought into existence wouldn't be benefited in the sense described in section I, but at least there would be people to point to who could at least in some sense be described as having benefited, whereas if they are not brought into existence, there is no such individual. In the negative utility case, for the same reason, one might think that there are individuals who have been harmed if the additional people are brought into existence, but that no one has been benefited if they are not brought into existence. Thus, through a quasi-person-affecting lens, the positive utility case involves either a benefit or nothing, and the negative utility case involves either a harm or nothing.

This observation alone, however, does not explain the intuition that numbers might matter in the negative utility case but not in the positive utility case. There still is a benefit in one case and a harm in the other, and the intuition we are trying to explain is that we are neutral about the additional benefit, but not neutral about the additional harm. An

https://doi.org/10.1017/S0953820815000072 Published online by Cambridge University Press

unexplained asymmetry remains. However, once we see that person-affecting intuitions probably influence us to view the positive utility case solely in terms of a benefit or no benefit and the negative utility case solely in terms of a harm or no harm, the heavy lifting has been finished. It is well known that our common-sense moral intuitions take into account distinctions such as those between doing and allowing and between conferring harms and failing to confer benefits. Thus, once viewed through this lens, it is unsurprising that even those who think population size is a matter of taste in the positive utility context will find it to be a moral matter – and not merely a matter of taste – in the negative utility context.

So far so good, but how does this support AU? This supports AU because it casts doubt on our intuitions regarding Two Hells and cases similar to Two Hells. First, it seems that the intuition against AU in cases like Two Hells is quite possibly a function of person-affecting intuitions, and it has already been shown why they are problematic in the population ethics context. *Further*, it seems that the intuitions on which they rely in this context are quite possibly ones rooted in distinctions – such as the distinction between harming and failure to benefit – that are components of deontological theory, and which not only are contestable in their own right, but at the very least seem inapplicable in our current inquiry.

Thus, these considerations might provide the beginnings of an explanation of why someone who is inclined to be in favour of AU in the positive utility context is perhaps less willing to be on board with AU in the Two Hells-type context.[52] Further, as argued in section VI.A, *even*

[52] Not only has the foregoing discussion attempted to provide an explanation of why one might have asymmetrical intuitions in the positive and negative utility contexts, but it has aimed to show that – once made aware of this explanation – one should revise one's view and maintain a position that has a symmetrical treatment of the two contexts. It is important to note, however, that even this – if accepted – is not (by itself) enough to show that we should reject our prima facie TU-friendly intuitions in the Two Hells context to make them consistent with AU-friendly intuitions in the positive utility context. One could consistently make the opposite adjustment instead. While I do not have enough space here to articulate my explanation fully, the following is my view in a nutshell. First, I believe I have provided strong reasons (throughout this article) to accept average utilitarianism in the positive utility context. If plausible, these provide strong reasons to maintain AU-friendly intuitions in the positive utility context and to revise one's TU-friendly intuitions in the Two Hells context – as opposed to making the opposite adjustment. Second, my explanation for the different intuitions in the positive utility context and the Two Hells context operates by showing how person-affecting intuitions are seeping in and how our person-affecting intuitions provide different responses in the context of benefits and in the context of harms. The difference, however, in brief, is that we care more about the effect *for a person* in the harm context than in the benefit context. Thus, since I have argued that person-affecting intuitions should be rejected, it is our intuitions in the context of benefits – and thus, in this case, the positive utility

https://doi.org/10.1017/S0953820815000072 Published online by Cambridge University Press

if people might be inclined to have anti-AU intuitions in the Two Hells case itself (and thus reject the arguments in this section), suffering cases *on the whole* provide no more support for TU than they do for AU. For these reasons, Parfit's Two Hells objection – at the absolute least – should not be considered a strong reason for rejecting AU.

Before closing, it is important to address a common misconception about an aspect of AU's treatment of the Two Hells case. One might ask why it wouldn't be a good thing to give the suffering souls what they want, i.e. death. AU would appear to say that the death of a suffering individual, assuming his happiness level is equal to the average in the population, would be neither good nor bad. While perhaps some versions of AU – namely AU2 or AU3[53] – would in fact imply this, the AU account endorsed in this article does not. In comparing either Hell 1 or Hell 2 to a state of affairs, Z, in which one of the individuals, A, dies, it is not the case that Z does not include A. Rather, in Z, A would exist and have a less negative value than in the original state of affairs, because he would have experienced less lifetime suffering. As such, on this AU account, the death of a suffering individual *would* be a good thing because it would increase the average happiness of the population. Further, and for the same reason, the death of all existing individuals would be an even better result.

CONCLUSION

A notable aspect of the AU account espoused in this article is that the total quantity of happiness does not matter on the population level, but the total quantity of happiness does matter on the individual level. In other words, all else equal, it is not better for there to be more happy lives than fewer happy lives, but, all else equal, it is better to have a longer happy life than to have a shorter happy life. I think that part of what has contributed to the widespread rejection of AU is the failure to recognize that this asymmetry is not implausible. Huemer argues, and others assume, that if quantity has value intrapersonally, then it must also have value interpersonally. In order to espouse AU, it seems that people assume that one must adopt an intrapersonal view according

context – that more closely approximate an impersonal morality, and, thus, it is these that we should maintain. For this reason, our intuitions in the Two Hells context are more suspect, and it is the intuitions in this context that should be rejected. I owe great thanks to an anonymous reviewer for pointing out the need to provide an explanation for why one should revise one's conflicting intuitions in the manner I suggest we should, as opposed to doing so in the opposite manner.

[53] See section IV.

https://doi.org/10.1017/S0953820815000072 Published online by Cambridge University Press

to which the total quantity of happiness does not matter. This would perhaps involve adopting an Epicurean view of death, or something similar – perhaps AU2 or AU3, as discussed in section II – and this is not a view that many people find attractive. But this choice is a false one. We can have the best of both worlds.[54]

mpressman@gmail.com

[54] I am deeply indebted to Ralph Wedgwood for his comments – comments that have been both numerous and extraordinarily helpful. I am also grateful to two anonymous reviewers for their valuable input.

https://doi.org/10.1017/S0953820815000072 Published online by Cambridge University Press