

On the Survival of Humanity

Johann Frick

*Department of Philosophy and Center for Human Values, Princeton University,
Princeton, NJ, USA*

Contact: jdfrick@princeton.edu

Abstract: What moral reasons, if any, do we have to ensure the long-term survival of humanity? This article contrastively explores two answers to this question: according to the first, we should ensure the survival of humanity because we have reason to maximize the number of happy lives that are ever lived, all else equal. According to the second, seeking to sustain humanity into the future is the appropriate response to the *final* value of humanity itself. Along the way, the article discusses various issues in population axiology, particularly the so-called Intuition of Neutrality and John Broome's "greediness objection" to this intuition.

Keywords: Population ethics; axiology; survival of humanity; Intuition of Neutrality; John Broome.

1. Introduction

It is probably inevitable that humanity will someday go out of existence. The end could be sudden and cataclysmic – the result of an asteroid strike or nuclear war. Or extinction might creep up on us gradually, through catastrophic climate change, resource depletion, or falling birthrates. What seems hard to doubt is that, one way or another, humanity's career will eventually come to an end. In the extremely unlikely event that mankind has managed to survive until then, the end of the Universe will finally do us in.

Most people accept that humanity's demise is only a matter of time. Yet, at the same time, many of us believe that it would be very bad, indeed one of the worst things that could possibly happen, if, for preventable reasons, the end came much sooner rather than later. Insofar as this is within our power, we assign significant importance to ensuring that humanity survives for as long as possible.

These sentiments can be explained as arising, in part, from self-interest. Most obviously, there is the small but non-trivial possibility that *we* will be part of that last generation which is violently extinguished. No-one wants to belong to the generation that allowed humanity to go extinct – if only because this might coincide with our own premature demise.

Less trivially, we may also have wider self-regarding reasons for wanting humanity to continue *beyond* our own deaths – for wanting there to be, in the words of Samuel Scheffler, an “afterlife” (Scheffler, 2013). Many of the activities and practices that give meaning to our own lives, Scheffler argues, are intimately connected with, and sustained by, our belief in the continued survival of humanity. In the absence of an afterlife, goal-oriented activities such as cancer research or social reform would lose much of their sense, since a large part of the intended benefit of these activities would vanish together with mankind itself. Likewise, activities designed to sustain certain values over time, such as traditions or cultural practices, would be threatened with futility, since the end of humanity would preclude the success of such efforts.

Yet, although the factors highlighted by these self-regarding accounts play an important role in our thinking about humanity's survival, they do not tell the full story, I

believe. Even a Schefflerian interest in the afterlife, which transcends a mere concern for our *personal* survival, can be expected to drop off rather quickly as the things that make us care about posterity – a “personalized” relation with the future or the survival of certain cherished values or traditions – fade over time. Yet many of our present efforts to control existential risks to humanity, such as early detection programs for large earth-bound asteroids, involve extinction-events that would, in all likelihood, occur many centuries, if not thousands of years, from now. If, indeed, there are good reasons to undertake these efforts, they cannot primarily be grounded in our collective self-interest.

Jonathan Glover (1977, 69) reaches a similar conclusion by means of a thought-experiment: he has us imagine that we could take a drug which causes everyone currently alive to become infertile, but at the same time makes us so blissfully happy as to become completely indifferent to this fact – thereby ensuring that the youngest generation alive will be the last. Even in this scenario where, Glover assumes, any *self-regarding* reasons to prefer the survival of humanity are outweighed or defeated, it may be wrong for us to let humanity become extinct. Like many others, Glover believes that mankind has weighty *moral* reasons to ensure that it survives for as long as possible.

The aim of this essay is to explore what, if anything, might ground such moral reasons.¹ The thought I wish to examine is that we have moral reasons to care for the

¹ My constructive proposal will not take a stand on how weighty these moral reasons are, nor on what, *exactly*, they are reasons to do. Before we can sensibly attack these issues, we must first resolve the more fundamental question whether there *are* any genuinely moral reasons to ensure the survival of humanity in the first place. This is the task I set myself in this article.

survival of humanity because this is, in a sense yet to be determined, the appropriate way to respond to the *value* of individual human lives, or of humanity as a whole.

I contrastively examine two arguments which spell out this thought: the *Argument from Additional Lives*, favored by many consequentialist philosophers, and a non-consequentialist alternative, the *Argument from the Final Value of Humanity*.

Let us turn, first, to the Argument from Additional Lives.

2. The Argument from Additional Lives

What moral reasons might there be to help ensure that humanity survives for as long as possible? By far the most common response found in the literature is what I shall call the *Argument from Additional Lives*, which is favored by many consequentialist philosophers.² The argument proceeds as follows:

The Argument from Additional Lives:

- (1) The world will go better, all else equal, if and because more happy lives are lived than if they are not lived.

² Early versions of this argument are found, inter alia, in Glover (1977), Kavka (1978), and Parfit (1984).

- (2) The world will go worse, all else equal, if and because more miserable lives are lived than if they are not lived.³
- (3) If humanity survives longer rather than going extinct sooner, the goodness of the additional happy lives that will be lived as a result will outweigh the badness of the additional miserable lives that will also be lived.
- (4) Hence, all else equal, the world goes better if humanity survives longer rather than going extinct sooner. [from (1), (2), and (3)]
- (5) We have a pro tanto moral reason to do what will make the world go better.
- Therefore, all else equal, we have a pro tanto moral reason to ensure that humanity survives longer rather than letting it go extinct sooner. [from (4) and (5)].⁴

The assumption in premise (3) – that if we prevent the premature extinction of humanity, this will increase the net balance of happy over miserable lives that are ever lived – is non-

³ In keeping with established practice in population ethics, I employ the phrases “a happy life” and “a miserable life” as synonyms for “a life worth living” and “a life not worth living”. In so doing, I do not take a stand on what, exactly, makes a life worth living or not worth living for a person. That is, I shall remain agnostic with regard to the correct theory of well-being (which I understand as that which makes a person’s life go well or, at least, worth living).

⁴ The qualification “all else equal” in premises (1), (2), (4) and the conclusion is necessary in order to account for the possibility that, despite making the world better in one respect, the longer survival of humanity makes the world worse *all things considered*, for instance by greatly increasing suffering among non-human animals. In the following, like most proponents of the Argument from Additional Lives, I shall bracket this possibility. I thank Krister Bykvist for discussion of this point.

trivial and might be contested. However, given likely advances in medicine, agriculture, and technology over the coming centuries, and in spite of serious challenges such as global warming, the gradual exhaustion of non-renewable resources, etc., it is a reasonable assumption to make. Let us grant it for the sake of argument.

Likewise, a skeptic might attempt to resist the Argument from Additional Lives by denying its fifth premise, the *deontic* claim that we have a pro tanto moral reason to do what will make the world go better. I have argued elsewhere that such attempts are unpromising.⁵ I shall not restate these arguments here.

The key moral premise of the Argument from Additional Lives, and the one I wish to focus on in the following, is proposition (1), the *axiological* claim that the world will go better just in virtue of additional happy lives being lived.

One canonical argument in support of this claim is provided by classical utilitarianism and other *totalist* views. According to this school of thought, the goodness of a state of affairs is some additive function of the wellbeing contained in individual lives. For a totalist utilitarian, adding an extra happy life to an existing population makes the world better, all things equal, since it increases the net total of pleasure minus pain that it contains.

⁵ See my doctoral dissertation “*Making People Happy, Not Making Happy People*”: *A Defense of the Asymmetry Intuition in Population Ethics* (Harvard University, 2014), Chapter 2.

But arguments in support of proposition (1) need not be this philosophically involved. Gregory Kavka states a simpler argument, which, for ease of future reference, I will dub the *Argument from the Value of Life*:

The lives of future people would almost certainly possess the properties that make the lives of present people valuable, and hence would be valuable themselves. This seems to be a reason for creating such lives; that is, for bringing future people into existence. (Kavka 1978, 195-6)

We will return to these arguments below.

For the time being, note that if the case for preventing the premature extinction of humanity is essentially that this will increase the number of happy lives that are ever lived, which makes the world go better, then it is plausible that our moral reasons for ensuring the survival of humanity are *temporally neutral*. That is to say, it is morally irrelevant that by enabling humanity to survive longer we cause additional lives to be lived in the *future*. This is because, according to most proponents of the Argument from Additional Lives, the contributory value of a happy life (i.e. its contribution to making the world go better) does not depend on *when* it is lived.⁶ According to these philosophers, our moral reasons for wanting humanity to become *diachronically* more numerous, by surviving for longer, are thus equivalent to our reasons for wanting it to be *synchronically* more numerous, by

⁶ For discussions of this point, see Glover (1977, 69-71) and Broome (2004, 126-128). See Footnote 11 for why, strictly speaking, Broome is not himself a proponent of the Argument from Additional Lives.

containing more people in any given generation. This equivalence is explicitly affirmed by Kavka:

Imagine God deciding between creating a universe with one planet occupied by n happy people, and a universe with two planets, *each* occupied by n people, each just as happy as those in the first universe. Does the fact that there are twice as many happy people in the latter universe constitute a reason for God preferring to create it? (...) Notice that our problem concerning future generations is quite analogous to this one, the difference being that the extra equally happy people are located in later generations rather than on another planet. Now I confess to being one of those who strongly feels it would be better for God to create the greater number of equally happy planets (or generations). (Kavka 1978, 196)

3. The Intuition of Neutrality

Many persons, myself included, are not compelled by the first premise of the Argument from Additional Lives. Instead of holding that each additional happy life makes the world better, many of us are inclined to believe that, setting aside its effects on other people, it is often *axiologically neutral* whether or not additional people are brought into existence. As John Broome puts it:

[T]he presence of an extra person in the world is neither good nor bad. More precisely: a world that contains an extra person is neither better nor worse than a world that does not contain her but is the same in other respects. (Broome 2005, 401).

Call this the *Intuition of Neutrality*. Proponents of the Intuition of Neutrality, of course, do not deny that were we to create a person whose life was so *bad* as to not be worth living, this would make the world worse, all things equal. Rather, the claim is that in the case of a

worthwhile extra life, “there is no consideration stemming from the wellbeing of the person herself that counts either for or against bringing her into existence.” (Broome 2004, 144)

Importantly, the Intuition of Neutrality assumes that there is a *range* of different levels of wellbeing at which adding a person to an existing population is ethically neutral, rather than a single neutral level. This neutral range has a lower bound (though it may not be sharp), below which adding a person to an existing population makes the outcome worse⁷; however, it may or may not be upwardly unbounded.⁸ (I myself am most attracted to a strong version of the intuition, on which the neutral range is upwardly unbounded).

Rejecting the Intuition of Neutrality, as the Argument from Additional Lives is committed to doing, comes with considerable costs:

⁷ This lower bound may correspond to the level of wellbeing at which a life becomes not worth living for that person herself; but on some views – so called *critical level* theories – the level of wellbeing below which adding an extra person to a population makes the outcome worse may be considerably higher than the boundary between a life worth living and a life not worth living. Broome (2004) defends a form of critical level theory.

⁸ If there is an upper bound to the neutral range, this implies that there is some threshold of personal value such that creating a new life above that threshold makes the world better. There are also other possibilities: if we are *perfectionists*, we may hold that there are some individuals whose existence makes the world better on account of their extraordinary achievements or because their lives are excellent in some other regard – irrespective of the amount of *wellbeing* that these lives contain. We may believe, for instance, that the existence of van Gogh or Kafka made the world better on account of their extraordinary artistic achievements – not because they had lives that went especially well *for them*, all things considered.

For starters, the Intuition of Neutrality neatly dovetails with a deontic intuition about the morality of procreation, the so-called *Procreation Asymmetry*, which strikes many as powerfully attractive. According to this intuition, if a future person would foreseeably have a life that is not worth living, this in itself gives us a strong moral reason to refrain from bringing this person into existence. By contrast, there is no moral reason to create a person whose life would foreseeably be worth living, just *because* her life would be worth living.⁹ If, on the other hand, we deny the Intuition of Neutrality and embrace the view that each additional happy life makes the world better all else equal, the Procreation Asymmetry may seem hard to defend. If creating a new happy life would make the world better, all else equal, shouldn't we have *some* moral reason to do so?

Rejecting the Intuition of Neutrality would have strongly revisionary implications in many other contexts, too. A nice example comes from Broome (2005, 402).

Considering the moral value of programs for improving people's safety, he writes:

When people's lives are saved, by making roads safer or in other ways, the wellbeing of the people who are saved is generally small in comparison to the wellbeing of all the new people, their descendants, who come into existence as a result. This is perfectly predictable. If all the descendants' wellbeing had to be counted too, that would enormously alter the value we attach to saving people's lives. But actually, in

⁹ The Procreation Asymmetry was first discussed by Narveson (1967). The label is due to McMahan (1981). I defend the Procreation Asymmetry at length in my article "Conditional Reasons and the Procreation Asymmetry" (unpublished manuscript).

judging the value of safety on the roads, we routinely ignore all this wellbeing.
(Broome 2005, p. 402)

Finally, rejecting the Intuition of Neutrality also has *theoretical* costs in the field of population ethics. Affirming that the world goes better, all else equal, if additional lives worth living are lived than if they are not lived invites a version of Parfit's *Repugnant Conclusion*; for it opens up the possibility that adding sufficiently many lives that are *barely* worth living to a world can morally outweigh a reduction in the well-being of an original population, in which everyone was very well off.

One contemporary philosopher who claims to be strongly attracted to the Intuition of Neutrality is John Broome himself. In a series of articles going back to the early 1990s and culminating in his treatise *Weighing Lives*, Broome time and again returns to the intuition, running through a number of possible interpretations. In the end, he reluctantly concludes that, despite its initial plausibility, the Intuition of Neutrality is ultimately untenable. There is no way of fleshing out the intuition that avoids unacceptable implications. If there is a neutral level of existence, Broome maintains, this must be a *single* (albeit *vague*) level of wellbeing. Any extra life above that level makes the world better. The upshot of Broome's arguments, if correct, would be dramatic. Not only would we forfeit one principled basis for resisting the Argument from Additional Lives. Moreover, as Broome's own road safety example suggests, we would also be forced to

radically revise our ethical thinking about any aspect of our behavior or policies that is likely to impact the number of people who will live in the future.¹⁰

In the following, I seek to show that Broome is too pessimistic in his conclusions. His main argument against the Intuition of Neutrality, the *Argument from Greediness*, fails.

4. Two Interpretations of the Intuition of Neutrality

As it stands, the Intuition of Neutrality is too vague to be useful. What precisely do we mean when we say that it is “axiologically neutral” to add a happy person to the world?

One way of interpreting this thought is that the state of affairs that results from adding a person in the neutral range is *equally good* as the state of affairs containing only

¹⁰ Nota bene: That Broome believes we must give up the Intuition of Neutrality does not mean that he himself endorses the Argument from Additional Lives. Broome stops short of endorsing this argument since he is, for the time being, agnostic about where the single neutral level of existence is located (Broome, 2005, 208). Without further work in population ethics, Broome believes, we are not at present in a position to know whether the absence of worthwhile lives due to the premature extinction of humanity would make the world worse or better (though Broome certainly inclines to the former belief). What we *can* know, contra the Intuition of Neutrality, is that the absence of such a large number of worthwhile lives is likely to be ethically *highly significant* – one way or the other. As Broome puts it in his recent book *Climate Matters*: “Because the intuition of neutrality is false (...) we cannot assume that all those absences are neutral in value. We must expect the absences to be either a good thing or a bad thing. Intuitively it seems most plausible that they are bad. We mostly feel that (...) extinction would be a very bad thing indeed. But (...) we still have a lot of work to do before we can be sure that is so.” (Broome 2012, 183).

the initial population, all else equal. However, on this interpretation, Broome argues, the Intuition of Neutrality is not tenable. Consider the following three states of affairs, R, S, and T. The numbers in brackets are vectors representing the individual levels of wellbeing of a hypothetical four-person population (in the following, we assume that all positive numbers in my examples indicate levels of welfare in the neutral range). The letter “ Ω ” represents non-existence.

$$R = (5, 3, 5, \Omega)$$

$$S = (5, 3, 5, 3)$$

$$T = (5, 3, 5, 5)$$

If adding a person in the neutral range to an existing population results in a state of affairs that is *equally good*, then R and S are equally good, and R and T are also equally good. By the transitivity of the “equally good as”-relation, it follows that S and T must be equally good as well. This, however, is an unacceptable conclusion. T is clearly better than S; this follows from Broome’s very plausible *Principle of Personal Good*, according to which, if two distributions A and B have the same population, and A is at least as good as B for each member of the population and better for some member of the population,

then A is better than B.¹¹ If we interpret neutrality as equality of goodness, we will thus derive a false conclusion. We must therefore abandon this interpretation.¹²

It is not hard to find an interpretation of the intuition that appears more promising. We are ready to accept, Broome maintains, that there are cases where none of our standard trichotomy of value relations (“better than”, “worse than” and “equally good as”) apply. Rather, the two things being compared may be *incommensurate* in value, by which Broome just means that “neither is better than the other, yet they are also not exactly equally good.” (Broome 2005, 407)¹³

We are prepared to encounter incommensurateness of value in contexts where the two things being compared are sufficiently different in nature that a *precise* comparison between their goodness seems impossible. A popular illustration is artistic greatness across different domains of creation. Suppose you were asked: “Who was the greater artist: William Shakespeare or J.S. Bach?” You may be inclined to respond: “Neither of

¹¹ Cf. Broome (2005, 120).

¹² For an ingenious recent attempt to defend the equality-interpretation of the intuition of neutrality against Broome’s critique, see Melinda Roberts, “The Neutrality Intuition” (unpublished manuscript).

¹³ Broome’s term “incommensurateness” is ambiguous between two readings:

Incommensurateness, as Broome defines it, states only the absence of any of the standard trichotomy of evaluative relations. This leaves two options: either the two items are *incomparable* in value, or they are connected by a fourth positive value relation, for example Ruth Chang’s (2002) notion of “parity”. None of Broome’s arguments *against* the incommensurateness interpretation seem to turn on adopting one reading rather than the other. By contrast, the case *for* the incommensurateness interpretation seems to me strongest when it is cashed out in terms of parity. So that is the reading I will assume in the following.

the two was a greater artist than the other.” But must this imply that Shakespeare and Bach were *exactly equally* great artists? Surely not. The two domains of creation in which they were active – literature and music – are too different to allow such precise comparisons. An upshot of this is that if we imagined an artist, call him “Shakespeare Minus”, who, by assumption, was a *somewhat* less great artist than Shakespeare, it does *not* follow that Shakespeare Minus was a less great artist than J.S. Bach. (We assume that amongst writers more precise comparisons of artistic greatness are possible than between writers and composers).

Broome accepts, *arguendo*, that two populations containing a different number of lives may often be incommensurate in value. In *Weighing Lives*, he states this incommensurateness interpretation of the Intuition of Neutrality as follows:

Suppose two distributions have the same population of people, except that an extra person exists in one who does not exist in the other. Suppose each person who exists in both distributions is equally as well off in one as she is in the other. Then there is some range of wellbeings (called ‘the neutral range’) such that, if the extra person’s wellbeing is within this range, the two distributions are incommensurate in value. (Broome 2004, 167)

If the incommensurateness interpretation of the Intuition of Neutrality were tenable, we could avoid the problem that arose for the “equally good as” interpretation. For, unlike “equally good as”, the “incommensurate with”-relation is not transitive. Even though Shakespeare’s artistic greatness is incommensurate with that of J.S. Bach, and Bach’s artistic greatness is incommensurate with that of Shakespeare Minus, it does not

follow that Shakespeare Minus is incommensurate with Shakespeare in terms of their artistic greatness. By assumption, Shakespeare is a greater artist than Shakespeare Minus. Likewise, comparing our three populations above, we could say that the expanded population S is incommensurate in goodness with the original population R; and, for the same reason, R is incommensurate with T. Nonetheless, T, which is better for the fourth person and worse for none, is clearly better than S.

What is more, the axiological judgments suggested by the incommensurateness-interpretation of the Intuition of Neutrality square neatly with our *deontic* intuitions about such cases. Thus, in choosing between R, S, or T (i) if R and S were the only two options, it would be morally permissible to bring about either R or S; (ii) if R and T were the only options, it would be permissible to bring about either R or T; but (iii) if R, S and T were all possible, then – while it would be permissible to bring about either R or T – it would be *prima facie* wrong to produce S. This, I believe, is because outcome S is “dominated” by outcome T, which is strictly better than S, whereas both S and T are incommensurate with R.

Understanding neutrality as incommensurateness of value thus avoids some of the problems of the equality interpretation, and accounts for our deontic judgments about this case. Yet, in the absence of a deeper theoretical motivation, the appeal to incommensurateness may still appear *ad hoc*. In the Shakespeare vs. Bach example, a judgment of incommensurateness seems justified by the fact that the qualitative differences between music and literature are sufficiently great to render *precise* comparisons of artistic greatness across these different domains impossible. But, *prima*

facie, it is not clear that a mere difference in *size* between populations makes for an important qualitative difference.¹⁴

Broome himself does not attempt to supply a deeper motivation for the thought that the Intuition of Neutrality might be best cashed out in terms of incommensurateness. But I believe I can.

5. Incommensurateness and the Conditional Value of Wellbeing

I want to argue that the Intuition of Neutrality rests on a deeper truth about the value of wellbeing.¹⁵

A common criticism of totalist utilitarianism is that it does not take persons sufficiently seriously. It views them as fungible receptacles for wellbeing, not as mattering *qua* individuals. By treating the moral significance of persons as *derivative* of their contribution to valuable states of affairs in which total wellbeing is maximized, it *reverses* what strikes most of us as the correct order of dependence. Human wellbeing matters because people matter – *not vice versa*.

From this basic insight, we can derive a novel argument in support of the Intuition of Neutrality. I do so by exploring an interesting parallel between the Intuition of Neutrality and an analogous set of claims about the value of promise-keeping: Most of us believe that, all else equal, it makes the world go worse to make a promise and then fail to

¹⁴ Cf. Broome (2004, 168).

¹⁵ In the following, I restate, in axiological terms, points which I develop at much greater length in “Conditional Reasons and the Procreation Asymmetry”.

keep it. (Compare: All else equal, it makes the world go worse to create a life that is not worth living). By contrast, all else equal, we do not think that making and keeping a promise, by itself, makes the world go better (or worse) than never making the promise in the first place. As Holly Smith put it: “keeping a promise does not seem to add any moral value to the world that must be taken into account when deciding whether to make that promise.” (Smith 1997, 183) (Compare: all else equal, it makes the world go neither better nor worse to create a new happy life from within the neutral range).

In the case of promising, the latter judgment is not hard to explain: While it is better to keep our promises than to break them, all else equal, the value of keeping our promises is *conditional* on the promise having been made. There is no unconditional value in (making and keeping) a promise. I want to suggest, likewise, that human wellbeing has contributory value, but this value is *conditional* on the *existence* of those to whom it accrues.¹⁶ While it matters greatly whether an individual life, or an existing population, is more or less happy, there is no unconditional contributory value in creating further happy lives. Adding wellbeing to the world by increasing the happiness of existing people makes

¹⁶ More precisely: The conditional value of wellbeing is conditional on existence, but it is not conditional on the *identity* of those who exist. This claim is necessary in order to avoid Parfit’s (1984) Non-Identity Problem. Thus, in the following same-number case

$$P = (5, 3, \Omega)$$

$$Q = (5, \Omega, 5)$$

we can say that distribution Q is impersonally better than P, since it does better by the people who exist in it, although the persons who exist in P and Q are not identical.

the world better. By contrast, adding new happy people to the world is axiologically neutral. And adding a person whose life is not worth living makes the world worse.

Creating a new person, on my view, is like giving rise to a new *challenge*. You can fail at the challenge (if the new person's life is not worth living) or succeed at it to various degrees (if their life is more or less worthwhile). But the fact that you would meet the challenge that you give rise to by creating the person is not itself a reason for creating the person. Meeting the challenge successfully has no *unconditional* value.

The view that the contributory value of wellbeing is conditional on existence supports the incommensurateness-interpretation of the Intuition of Neutrality. Since wellbeing lacks unconditional value, adding a new person whose wellbeing falls within the neutral range makes the outcome neither better nor worse compared to an outcome that doesn't contain this person. Moreover, this will be true across a *range* of levels of wellbeing, since, compared to creating no new person, neither creating a very happy new person nor creating a moderately happy person has unconditional contributory value. On the other hand, once we fix the fact that a new person will exist, it is strictly *better* that her life contain more wellbeing rather than less. As we saw above, these are exactly the judgments implied by the incommensurateness-interpretation of the Intuition of Neutrality.

6. Broome's Argument from Greediness

Unfortunately, Broome believes, even the incommensurateness interpretation of neutrality

is ultimately untenable. It falls prey to what I will call his *Argument from Greediness*.¹⁷

Consider the following three distributions:

$$A = (3, 2, 5, \Omega)$$

$$B = (3, 2, 5, 1)$$

$$C = (3, 2, 4, 4)$$

We stipulate that B is produced from A by adding a person within the neutral range. In line with the incommensurateness interpretation of neutrality, we suppose that A and B are incommensurate in value. Now compare B and C. Both these distributions contain the same people. The difference between them is that in C one person is made worse off and one person is better off than in B. Broome asks us to assume that, all things considered, C is better than B, which appears plausible: The third (better-off) person's wellbeing has diminished by less than the fourth (worse-off) person's wellbeing has increased; also, C is a more equal distribution than B. Anyone with egalitarian sympathies should agree that C is better than B. Since B is not worse than A, Broome concludes, it follows that C is not worse than A (since, in general, if y is not worse than x, and z is better than y, then z is not worse than x). Call this the *Neutralist's Argument*, for ease of future reference.

¹⁷ I here follow, step for step, Broome's presentation of the argument in his article "Should We Value Population?", *The Journal of Political Philosophy*, Vol. 13, No. 4 (2005), in particular pp. 407-411. A shorter statement of the Argument from Greediness occurs in *Weighing Lives*, pp. 169-70.

But this conclusion, Broome now argues, is unacceptable. Compare A and C: These two distributions differ in two respects: First, one person is worse off in C than A. In this respect, C is unequivocally worse than A. Second, C contains an extra person, whose addition, we assumed, was axiologically neutral. In this respect, Broome claims, C is neither better nor worse than A. So C is worse than A in one respect, and neither better nor worse in the other respect. Intuitively, therefore, C must on balance be worse than A. As Broome puts it: “In going from A to C, we have one bad thing and one neutral thing. A bad thing plus a neutral thing must add up to a bad thing.” (Broome 2005, 409).

However, the intuition of neutrality has led us to the contrary conclusion, namely that C is *not* worse than A. In going from A to C – although adding the fourth person is claimed to be axiologically neutral – this has resulted in the badness of the harm done to the third person being ‘cancelled out’. As Broome writes:

We have found that our neutrality is greedy. Although neutral in itself, it is able to swallow up bad things and neutralize them. (...) Intuitively, neutrality should not behave like that; it should not be greedy. (Broome 2005, 409)

For this reason, Broome believes, we must reject the incommensurateness interpretation of the intuition of neutrality.¹⁸ Since no other plausible interpretation of the intuition suggest

¹⁸ Broome (2004) also advances a second argument against incommensurateness interpretation, to wit that the boundaries of the neutral range must surely be *vague*, but that such vagueness at the boundaries is incompatible with the intuition of neutrality understood as

itself, Broome claims that we must reject the intuition outright.

This, I believe, is a mistake. I will now argue that we should reject Broome's Argument from Greediness, because it rests on an understanding of neutrality that is not only morally implausible, but also contradicts a principle that Broome himself affirms.

7. Rebutting Broome's Argument from Greediness

Broome argues that, in order to avoid being misled by the greediness of neutrality, the correct way for someone attracted to the Intuition of Neutrality to analyze the move from A to C is as follows: First, we compare only the effects on the three people who already existed in A. The move here is from $A = (3, 2, 5)$ to $C_1 = (3, 2, 4)$, which is a change for the worse. (From the Principle of Personal Good) Second, we consider the effect of adding the fourth person. The move here is from $C_1 = (3, 2, 4)$ to $C = (3, 2, 4, 4)$, which (by assumption) is axiologically neutral. Since one of these moves is worse, and the other is neutral, Broome concludes that the total effect of these two moves must be to make the distribution worse. The problem with the incommensurateness interpretation of the intuition is that it seems committed to denying this.

Broome's Argument from Greediness implicitly posits a constraint that any interpretation of neutrality must satisfy in order to be intuitively acceptable. Let us call this Broome's

incommensurability. This argument is convincingly rebutted by Rabinowicz (2009), so I will not review it here.

Non-Greediness Principle: If O_1 and O_2 are two outcomes, where O_1 contains an original population $P = (p_1, p_2, \dots, p_n)$ and O_2 contains P , plus a number of new individuals $N = (n_1, n_2, \dots, n_n)$ whose welfare levels all lie within the neutral range, then if O_2 is all-things-considered worse for the members of P than O_1 , then O_2 as a whole is worse than O_1 , and if O_2 is all-things-considered better for the members of P than O_1 , then O_2 as a whole is better than O_1 .

Only a conception of neutrality which does not violate this Non-Greediness Principle is one on which the addition of new persons to a population cannot ‘swallow up and cancel out’ bad or good things that happen to members of an original population.

By contrast, a conception of neutrality is greedy just in case it violates the Non-Greediness Principle, that is, just in case it accepts:

Greediness: If O_1 and O_2 are two outcomes, where O_1 contains an original population P and O_2 contains P plus a number of new individuals N whose welfare levels all lie within the neutral range, then O_2 may not be worse (better) than O_1 , despite being all-things-considered worse (better) for the members of P .

The incommensurateness reading of neutrality is greedy because, as we saw, it implies that population C in our example above is not worse than the original population A , despite being all-things-considered worse for those individuals who already existed in A .¹⁹

¹⁹ The same example could also be used to show that the equality interpretation of neutrality, which we discussed above, is likewise greedy. I leave this as an exercise for the reader.

What remains to be seen, however, is whether greediness is indeed an unacceptable property for a conception of neutrality to possess, as Broome claims.

In his discussion of Broome's Argument from Greediness, Wlodek Rabinowicz correctly points out that, as Broome originally characterized the Intuition of Neutrality, there is no *contradiction* between neutrality and greediness: "That adding people is (axiologically) neutral simply means that it on its own makes the world neither better nor worse." (Rabinowicz 2009, 399) However, "there is no reason to expect that adding things that are neutral in *this* sense will have no neutralizing effects on bad or good things that are being added at the same time. (...) [Hence] there needn't be anything wrong with greedy neutrality, if neutrality is interpreted as it was supposed to be interpreted." (Rabinowicz 2009, 399)

This rejoinder, however, seems insufficient to conclusively answer the Argument from Greediness. Even if the Non-Greediness Principle is not contained in Broome's statement of the Intuition of Neutrality itself, it could nonetheless be a plausible *constraint* on any acceptable interpretation of this intuition. To put the Argument from Greediness to rest once and for all, this is the claim we must challenge.²⁰

²⁰ In his most recent discussion of the Intuition of Neutrality in *Climate Matters*, Broome – perhaps in response to Rabinowicz's critique – appears to *build* the Non-Greediness Principle *into* his statement of the Intuition of Neutrality itself. Thus, he writes: "Suppose two alternative options A and B have just the same population of people, except that there are some people in B who do not exist in A. Call the people who exist in both A and B "the existing people" and others "the added people". The intuition [of neutrality] is that there is a neutral range of well-being such that, provided the added people's well-being is within the neutral range, the

I will adduce three considerations which show that the Non-Greediness Principle is not, in fact, a plausible constraint on the Intuition of Neutrality. First, I will suggest that a general principle that seems to undergird Broome's Argument from Greediness is not in general valid. Second, I will demonstrate that the Non-Greediness Principle is incompatible with a highly plausible axiological principle, the *Principle of Impartiality*, to which Broome himself is committed. Third, I will argue, contra Broome, that the morally most plausible specification of the Intuition of Neutrality is one on which neutrality *is* greedy.

Broome's argument from Greediness seems to draw its intuitive support from a more general principle, quoted above, namely that "a bad thing plus a neutral thing must add up to a bad thing." (Broome 2005, 409). Stated a bit more precisely, the idea seems to be that if an overall change, such as the move from A to C, can be decomposed into (i) a change for the worse – such as the move from A to C_1 – plus (ii) a move between two things that are incommensurate in value – such as the move from C_1 to C –, it must overall

following is true: if B is better than A for the existing people, then B is better than A, and if B is worse than A for the existing people, then B is worse than A." (Broome 2012, 176). This characterization of the Intuition of Neutrality, which includes the Non-Greediness Principle by *default*, is a significant departure from Broome's earlier statements of the intuition, such as the one I quoted in Section 3. According to that earlier statement, it is a sufficient *but not a necessary* condition for the move from O_1 to O_2 to be axiologically neutral that O_2 is neither better nor worse in its effects on the original people. Since I will go on to argue, contra Broome, that the most plausible specification of the Intuition of Neutrality does *not* hold it hostage to the Non-Greediness Principle, it is better to work with Broome's earlier statement of the intuition, which does not prejudge the issue by definitional fiat.

be a change for the worse. Call this the *Decomposition Principle*. The Non-Greediness Principle operationalizes the Decomposition Principle for the specific context of adding new people to an existing population.

The Decomposition Principle, however, does not appear to be valid in general. Consider another stock example from the literature on parity/incommensurateness, namely ordered pairs of careers and salaries (such as having a career as a painter, earning \$60k a year). A claim that is commonly made in this literature, and which strikes me as plausible, is that for many such ordered pairs, neither of two ordered pairs is better than the other, yet the two pairs are also not exactly equally good, on account of the qualitative differences between the careers in question.²¹ Concretely, let us assume, for the sake of argument, that for me a career as a painter, earning \$60k a year, is incommensurate in value with a career as a musician, also earning \$60k a year. By contrast, a career as a painter in which I earn only *slightly* more, say \$60.5k, is *better* than a career as a painter in which I make \$60k. Suppose, then, that starting off as a painter earning \$60.5k, I make a career change, ending up as a musician earning \$60k (you might imagine, further, that there is a brief intermediate stage, at which I am a painter earning \$60k). This overall change can be decomposed into (i) a drop in salary, holding constant my career as a painter, and (ii) a change in career, from painter to musician, holding constant my salary of \$60k. The former is a change for the worse whereas the latter, we assumed, is a move between two things that are incommensurate in value. According to the Decomposition

²¹ The locus classicus is Chang (2002).

Principle, my career change must therefore be a change for the worse, overall. But, intuitively, this is false. My new career as a musician, earning \$60k a year, may well be incommensurate in value with my previous career as a painter, earning \$60.5k a year.²²

That the Decomposition Principle fails to hold in general should make us suspicious about the Argument from Greediness. But this doesn't yet establish that the more specific and restricted Principle of Non-Greediness is not, as Broome believes, a constraint on any acceptable interpretation of the Intuition of Neutrality. It is to this task that I turn now.

Friends of the Intuition of Neutrality should, I believe, reject the Non-Greediness Principle, because it turns out to be incompatible with a highly plausible axiological principle, the *Principle of Impartiality*. This principle holds that, if two outcomes are described by vectors that are mere permutations of each other, the outcomes are equally good, all else equal. That is, outcomes:

$$X = (1, 2, 3, 4)$$

and

$$Y = (4, 3, 2, 1)$$

are equally good, all else equal, since the distribution of well-being in Y is merely a permutation of that in X. The principle is intuitively compelling. It captures a very attractive idea of ethical impartiality, namely that as far as the axiological evaluation of

²² I thank Krister Bykvist and Luc Bovens for helpful conversations on these points.

states of affairs is concerned, the well-being of all persons matters equally. Consequently, merely shuffling around well-being levels amongst a group of persons should result in an outcome that is equally good. In *Weighing Lives*, Broome himself explicitly affirms this principle.²³

To see why the Non-Greediness Principle contradicts the Principle of Impartiality, consider the following variation on our example from Section 6:

$$A = (3, 2, 5, \Omega)$$

$$C = (3, 2, 4, 4)$$

$$D = (3, 4, 4, 2)$$

Consider again the move from A to C. The Non-Greediness Principle implies that this must be a change for the worse, since C is all-things-considered worse for the three people who already existed in A (from the Principle of Personal Good). The addition of the fourth person cannot alter this conclusion, on pain of greediness. Hence, A is better than C.

Next, consider the move from A to D. Various plausible ethical views converge on the verdict that, as far as the effect on the first three persons is concerned, the shift from A to D is an improvement: total wellbeing has increased; the second (better-off) person has gained more than the third (worse-off) person has lost; and the distribution is more equal.

²³ Cf. Broome (2004, 135).

Hence, the Non-Greediness Principle implies that D must be all-things-considered better than A. Again, the addition of the fourth person cannot alter this conclusion, on pain of greediness.

The Non-Greediness Principle thus commits us to the view that D is better than A and that A is better than C. It follows from the transitivity of “better than” that D is better than C.

But now notice that the distributions $C = (3, 2, 4, 4)$ and $D = (3, 4, 4, 2)$ are, in fact, mere permutations of one another. According to the Principle of Impartiality, C is hence equally good as D. The Non-Greediness Principle thus leads to a contradiction with the Principle of Impartiality: If C is worse than A, and D is better than A, then C cannot be equally as good as D.

In a choice between Non-Greediness and the compelling Principle of Impartiality, is clearly the former that must go. The question is: must we give up the Intuition of Neutrality *along with* the Non-Greediness Principle? I will now argue that – far from being inseparable from the Non-Greediness Principle – the Intuition of Neutrality is actually *more* plausible if we free it from the shackles of this constraint.

The Non-Greediness Principle, in essence, amounted to giving *zero* weight to the wellbeing of the new persons n_1, n_2, \dots, n_n in O_2 , in the sense that even greatly improving their wellbeing at the expense of persons in the original population would necessarily result in the outcome becoming worse. This extreme bias towards previously existing persons is morally implausible. Once a person exists, her wellbeing matters equally to the goodness of a state of affairs, *even if* her initial creation was axiologically neutral. In

deontic terms, the rights and claims of existing people may impose constraints on the ways in which we may expand an existing population, and on the sacrifices that we may impose on presently existing people for the sake of as yet contingent future persons. By contrast, I claim that the axiological perspective ought not to display the same kind of special consideration for previously existing individuals. From the axiological perspective, in comparing two outcomes O_1 and O_2 , it ought to make no difference whether there are persons who exist both in O_1 or O_2 , or whether the two populations are numerically entirely distinct.

If this is right, then there is no reason to accept that outcome C is necessarily worse than A, as the Non-Greediness Principle would have it, just because it is worse for the people who already existed in A. Instead, we can agree with the Neutralist's Argument that, since C is better than B (in virtue of the significant improvements for the new person), and B is not worse than A (by the Intuition of Neutrality), C is not worse than A.

In general, I believe that the intuition of neutrality, as most people hold it, is better captured by the following

Principle of Holistic Neutrality: For any existing original population $P = (p_1, p_2, \dots, p_n)$ there may exist a *range* of levels of wellbeing such that the addition of some further person(s) X from that neutral range makes the resulting

population N neither better nor worse than, nor equally good as, P. This can be true *even if*, by adding X we also make some members of P worse off.²⁴

Unlike the Principle of Non-Greediness, the Principle of Holistic Neutrality gives non-zero moral weight to the level of welfare that X enjoys in N, even when this comes at the expense of members of the original population P. It is a morally more plausible specification of the Intuition of Neutrality.

8. What if the Intuition of Neutrality Were True?

I do not, by any means, take myself to have *established* the correctness of the Intuition of Neutrality. But hopefully I have done enough to illustrate its appeal, while also defending it against some prominent objections, to warrant investigating what would follow if the intuition were indeed correct.

If the Intuition of Neutrality were correct, this would give us grounds to reject the Argument from Additional Lives. Moral reasons, let alone an obligation, to ensure the survival of humanity could not be inferred from the contributory value of additional

²⁴ This principle does not commit me to the claim that *all* populations of different sizes are incommensurate in value. Laying out a theory that specifies exactly when, and on what grounds, a larger population is better or worse than a smaller population is beyond the scope of this article. My concern here has been merely to show that, for a large number of situations and for a large range of levels of wellbeing, adding worthwhile lives to an existing population may be axiologically neutral. This is all that is needed to undermine the Argument from Additional Lives.

worthwhile lives being lived, since in many cases the existence of additional worthwhile lives would have no contributory value, setting aside their effects on other people.

This conclusion might be thought to raise a worry: As Kavka's Argument from the Value of Life stressed²⁵, the lives of potential future persons would possess much the same properties as the lives of presently existing people. Now, if possessing these sorts of properties does not make it the case that the existence of these potential future persons would have positive contributory value, then what does that say about the value of presently existing people – indeed about the value of human life as such? Must we conclude that human lives are themselves 'without value', and that it doesn't matter whether our lives are worthwhile? By no means. Rather, I submit, Kavka's argument rests on a questionable assumption about what it is to respond appropriately to the value of human life.

Our conviction that individual lives and the wellbeing of individual human beings clearly *are* of value is evinced by many moral judgments we affirm: We condemn wars and genocide for leading to the large-scale destruction of human life. We criticize politicians or generals for "gambling with lives", and mean that they are failing to respect the value inherent in each human life. In general, we regard it as evident that we have moral reasons to care that the lives people lead go as well as possible. Where proponents of the Argument from Additional Lives go wrong is by assuming that – besides wishing to *protect*, to *respect*, and to *ameliorate* human lives – it is part of responding appropriately

²⁵ See Section 2.

to the value of human life that we must also seek to *promote* it, i.e. to want there to be as many worthwhile lives as possible.²⁶ This, I believe, is implausible. Borrowing Jan Narveson's (1967) dictum, I believe that what we should care about is to make people as happy as possible, not to make as many happy people as possible. It is for this reason that I incline to reject the Argument from Additional Lives. If we do have moral reasons to ensure that humanity survives for as long as possible, they do not reside in the goodness of maximizing the number of worthwhile lives that are ever lived.

9. The Argument from the Final Value of Humanity

Having set aside the Argument from Additional Lives, I will briefly sketch an alternative proposal, the *Argument from the Final Value of Humanity*. Doing so will allow us to build on the insights of Section 8 concerning the ways in which we appropriately respond to value.

Let us say that something has non-instrumental or, as I shall say, *final* value if it is valuable for its own sake, independently of the instrumental value it may have for anyone. In many cases, something with final value will be *intrinsically* valuable, i.e. valuable in virtue of its internal or non-relational properties; but it need not be so.²⁷

²⁶ For an illuminating discussion of values that we respond to appropriately by promoting and values that we respond to appropriately by respecting and protecting their embodiments, see Scanlon (1998, Ch. 2).

²⁷ For the distinction between intrinsic vs. extrinsic and instrumental vs. non-instrumental value, see Korsgaard (1996).

It is commonplace to claim of a wide range of things that they have final value in this sense: wonders of nature, great works of art, animal and plant species, languages, culture, etc. The suggestion that *humanity* too, with its unique capacities for complex language use and rational thought, its sensitivity to moral reasons, its ability to produce and appreciate art, music and scientific knowledge, its sense of history, and so on, should be deemed to possess final value, therefore strikes me as extremely plausible. I do not, however, have the space to argue this claim in this article.²⁸ I will ask you to grant it to me as a premise, in the interest of seeing whether the final value of humanity may ground a moral reason to ensure humanity's survival. What I shall argue in the following is that there is a link between responding *appropriately* to the final value of humanity and being at least disposed to ensure its survival.

The thought is, indeed, a natural one. We undertake great efforts to ensure the survival of many other things we consider finally valuable. Millions are spent annually on the preservation of rare plant and animal species (even those of no identifiable scientific or aesthetic value to us) and there are countless organizations worldwide working to preserve dwindling cultures and languages. (Again, the thought is not, necessarily, that being a member of this particular culture, or speaking that language, is *better for* anyone, nor indeed, that it is *better for* anyone, all things considered that there *exist* members of this culture or speakers of that language – though these things might also be true. Rather, it is held, the culture or language deserves to be preserved because they are valuable in their

²⁸ The claim that the human (and other) species have *intrinsic* value is defended at length in Rolston (1988).

own right.) The analogous thought that, given humanity's final value, we have a reason to care that it survives for as long as possible, is therefore hardly unmotivated.

Indeed, many philosophers have argued that caring for a thing's survival is, at least in normal circumstances, part of what it *means* to value that thing finally. Thus, Samuel Scheffler (2007, 106) writes: "What would it mean to value things but, in general, to see no reason of any kind to sustain them or retain them or preserve them or extend them into the future?"²⁹ G.A. Cohen, in his article "The Truth in Conservatism", makes the following argument: "A thing that has [final] value is worthy of being revered or cherished. We do not regard something as being worthy of being revered or cherished if we have no reason to regret its destruction, as such." This, Cohen thinks, is precisely what distinguishes valuing something finally from valuing it merely for its instrumental properties:

One can say, quite properly, upon acquiring a valuable thing, 'I shall value this until something better comes along', but one cannot in the same way say 'I shall cherish this until something better comes along': that could happen to be a correct prediction, but it could not express a decision to cherish. (Cohen, 20)

²⁹ So much at least seems true of things – like persons, objects, species, languages, etc. – which, given adequate care, *can* survive over extended periods of time. As Christine Korsgaard pointed out to me in conversation, the same concern for a thing's survival isn't necessarily part of responding appropriately to the value of essentially *ephemeral* things like sensory experiences or events (e.g. a concert). The relevant contrast in the appropriate attitudes may be roughly that between "cherishing" and "savoring".

Here we have then, I believe, a second way of explaining why we ought to care about humanity's survival, quite distinct from the Argument from Additional Lives: According to the *Argument from the Final Value of Humanity* (or *Argument from Final Value*, for short), each successive generation collectively has a pro tanto moral reason to work for the survival of humanity, since this is how we appropriately respond to the final value of humanity. Notice that unlike the Argument from the Additional Lives, the Argument from Final Value is not temporally neutral. What matters according to this argument is that humanity *survives*, i.e. that it persists into the future. The moral grounds that the argument advances for ensuring the survival of mankind, in part by helping future generations to come into existence, do not *eo ipso* constitute reasons for making humanity *synchronically* more numerous.

That is not to say that the Argument from Final Value is without implications for how many people there should be in any given human generation. In guarding against existential threats to humanity's long-term survival, it may be optimal for the size of the human population to fall within a certain range at any given moment (though that figure will certainly vary over time, depending on humanity's level of technological advancement, the earth's carrying capacity at that time, the array of existential threats we face, etc.). Moreover, to the extent that responding appropriately to humanity's final value gives us a reason to ensure, not that humanity grimly soldiers on, but that it survives in a *flourishing* state, this may give us further grounds for wanting a certain minimum population size: all else equal, a larger population is often more culturally dynamic and rich, the pace of technological and scientific progress is swifter, etc. The thought is merely

that, subject to satisfying these desiderata, which have their source in the final value of humanity itself, there is, under the Argument from Final Value, no *further* reason to keep adding more people to the population *just because* these people would have lives worth living.

We can further clarify our understanding of how the Argument from Final Value differs from the Argument for Additional Lives by addressing two objections from James Lenman's essay "On Becoming Extinct".

Like myself, Lenman, considers it "a natural thought (...) that the existence of human beings has [final] value, impersonally regarded. And that therefore it is a good thing that human beings should continue to exist for as long as possible." (Lenman 2002, 255)

Lenman, however, is skeptical of this argument. In particular, he worries that the implications of the Argument from Final Value might differ insufficiently from those of the Argument from Additional Lives, which he deems implausible. Lenman considers the example of a biological species that we might consider finally valuable: the white rhinoceros. He argues that if the final value of the white rhinoceros is a reason for wanting the species to survive for longer, then surely this must also be an argument for wanting it to be synchronically more numerous. Since Lenman believes this to be absurd, this supposed implication functions as a *reductio* of the Argument from Final Value: "If it is unclear how it would make things better to stretch out, synchronically, in a single generation, the numbers of white rhinos, it is unclear why it should make things better to stretch them out diachronically by having more generations." (Lenman 2002, 256)

Lenman's objection can be met. Like proponents of the Argument from Additional Lives, Lenman appears to assume that the only appropriate response to something's being valuable is to *promote* it, by increasing its incidence in the world. But this is a mistake. Cherishing, protecting, respecting, savoring, etc., are all modes of responding to something's final value which do not commit one to wanting to have more instances of it. Consider the question of *cultural* survival: It doubtless makes sense to maintain that aboriginal culture in Australia is finally valuable and should be helped to survive. But this does not commit one to claiming that the world would be better if aboriginal culture had *as many* members as possible. When what is finally valuable is a form of life or a species, what we ought to care about, we might say, is the ongoing instantiation of the universal, not the number of instantiations.

Note the importance of the word "ongoing" in the previous sentence. If I had said "what we ought to care about is the instantiation of the universal" this might have suggested the following argument: "If some thing x is finally valuable then, for any time t , it is impersonally better that there exist instances of x at t ."

This is a bad argument, as Lenman correctly points out: "We may think it a wonderful thing that the world contains many examples of jazz music, but how much should we regret its absence from, say, the world in the sixteenth century?" (Lenman 2002, 257). Likewise, we don't bemoan the fact that humans beings didn't exist in the age of the dinosaurs. Just as holding humanity to be finally valuable does not give us a reason to want there to be as many humans as possible at present, humanity's final value doesn't give us a reason to want it to be maximally extended across time. However, it is a mistake

to treat this as a *reductio* of our concern for the *future* survival of humanity, as Lenman does. This conclusion would only follow if our reasons for wanting there to be humans in the future were at the same time reasons for wanting humanity to have existed in the pre-historic past. And this, again, would only be true on a conception of final value as something to be promoted. On the understanding of final value that animates the Argument from the Final Value of Humanity, our reasons for wanting humans to exist in the future are not temporally neutral. Rather, they stem from our concern for the *survival* of those things we value finally.

10. Conclusion

If the tentative arguments of the preceding section are sound, I have identified a new basis for the claim that we have moral reasons to ensure that humanity survives for as long as possible. These reasons are grounded, not in the value of maximizing the number of worthwhile human lives that are lived, but in the final value of humanity itself, which gives us moral reasons to cherish and preserve it.

At the same time, my discussion leaves many open questions. I shall close by acknowledging two of them: First, as I stated at the outset of this essay, I do not have a worked-out view on how strong our moral reasons to ensure the survival of humanity are, nor on what sacrifices those alive at any given point in time could be required to make in pursuit of this aim (or on how these sacrifices would be distributed equitably). What I do believe is that any moral considerations requiring us to help ensure the survival of humanity must, in principle, be defeasible. We can imagine circumstances where

conditions on Earth (or on other planets) become so bleak for the foreseeable future as to make the lives of our immediate descendants not worth living. (Picture conditions after a nuclear holocaust). In such a situation, any moral reason to perpetuate the human race would presumably be outweighed or cancelled, or else bump up against moral constraints against creating children with lives that are not worth living. This could be true, even if humanity's *long-term* prospects – ten generations from now, after the surface of the earth has once again become inhabitable – were quite good.

Second, my discussion leaves open what *exactly* would count as the “survival of humanity”. What seems clear is that our reasons to ensure the survival of *humanity* go beyond merely making it the case that *human beings exist in the future*. Imagine a world in which each generation of humans dies and vanishes without trace before the next one is born (perhaps, like mayflies, each generation of human lays eggs before its death, but disappears before their offspring has hatched). Each new generation lives without knowledge of previous generations of humans. The human *species* survives in this scenario, but a lot of what we mean by “humanity”, and a lot of what seems uniquely valuable about it – our sense of history, cultural traditions, relationships between parents and children, etc. – is lost.

What about the possibility, advocated by ‘transhumanist’ philosophers such as Bostrom (2008), that by deliberately modifying and enhancing its biological constitution, *homo sapiens* might one day evolve to a ‘posthuman’ stage: Would this transition mean the end of “humanity”, and, if so, would it involve any kind of loss? These are subtle taxonomical and normative questions that I must leave for another occasion.

Acknowledgments

For their generous help and feedback on earlier drafts of this article, I would like to thank Nick Beckstead, Luc Bovens, Krister Bykvist, Ruth Chang, Partha Dasgupta, Louis-Philippe Hodgson, Tom Hurka, Mark Johnston, Frances Kamm, Christine Korsgaard, Eden Lin, Elinor Mason, Kian Mintz-Woo, Ekédi Mpondo-Dika, Derek Parfit, Philip Pettit, Theron Pummer, Mozaffar Qizilbash, Melinda Roberts, Geoff Sayre-McCord, Alison McQueen, Tim Scanlon, Alec Walen and Monique Wonderly, as well as audiences and workshop participants at Harvard University, the University of Oxford, Princeton University, and the University of Toronto. Finally, I would like to thank my hosts at the University of Toronto Centre for Ethics, where I completed this article.

Notes on contributor

Johann Frick is assistant professor in the Department of Philosophy and the Center for Human Values at Princeton University. His primary research interests lie in moral philosophy, political philosophy, and bioethics. His current work focuses on population ethics, the concept of interpersonal justification, and the ethics of risk imposition.

References

- Bostrom, N., “Why I Want to Be a Posthuman When I Grow Up’ in B. Gordijn and R. Chadwick (eds.) *Medical Enhancement and Posthumanity* (Springer, 2008), pp. 107-137.
- Bradley, B., “The Value of Endangered Species”, *The Journal of Value Inquiry*, Vol. 35 (2001), pp. 43-58.
- Broome, J., *Climate Matters: Ethics for a Warming World* (New York: Norton, 2012).

----, "Should We Value Population?", *The Journal of Political Philosophy*, Vol. 13, No. 4 (2005), pp. 399-413.

----, *Weighing Lives* (Oxford: Oxford University Press, 2004).

----, "The Value of Life" in his *Ethics out of Economics* (Cambridge: Cambridge University Press, 1999).

Chang, R., "The Possibility of Parity", *Ethics*, Vol. 112 (July 2002), pp. 659-688.

---, "Introduction" in R. Chang (ed.) *Incommensurability, Incomparability, and Practical Reason* (Cambridge, MA: Harvard University Press, 1997).

Cohen, G.A., "A Truth in Conservatism: Rescuing Conservatism from the Conservatives" (unpublished draft).

Frick, Johann, "Conditional Reasons and the Procreation Asymmetry" (unpublished manuscript).

---, "*Making People Happy, Not Making Happy People*": *A Defense of the Asymmetry Intuition in Population Ethics* (Doctoral dissertation, Harvard University, 2014).

Glover, J., *Causing Death and Saving Lives* (Harmondsworth: Penguin Books, 1977).

Kavka, G., "The Futurity Problem" in R. I. Sikora and B. Barry (eds.), *Obligations to Future Generations* (Temple University Press, 1978)

Korsgaard, C., "Two Distinctions in Goodness" in her *Creating the Kingdom of Ends* (Cambridge: Cambridge University Press, 1996).

Lenman, J., "On Becoming Extinct", *Pacific Philosophical Quarterly*, Vol. 83 (2002), pp. 253-269.

McMahan, J., "Review: Problems of Population Theory", *Ethics*, Vol. 92, No. 1 (1981), 96-127.

Narveson, J., "Utilitarianism and New Generations", *Mind* 76 (1967), pp. 62–72.

Parfit, D., *Reasons and Persons* (Oxford: Oxford University Press, 1984).

Rabinowicz, W., "Broome and the Intuition of Neutrality", *Philosophical Issues* 19 (2009), pp. 389-411.

Roberts, M., "The Neutrality Intuition" (unpublished manuscript).

Rolston, H., "Duties to Endangered Species", *Bioscience*, Vol. 35, No. 11, The Biological Diversity Crisis. (Dec., 1985), pp. 718-726.

Scanlon, T., *What We Owe To Each Other* (Cambridge: Belknap Press, 1998).

Scheffler, S., *Death and the Afterlife* (Oxford University Press, 2013).

---, "Immigration and the Significance of Culture", *Philosophy & Public Affairs* 35, No. 2 (2007), 93-125.