

On the Wrongness of Human Extinction

Simon Beard and Patrick Kaczmarek

Centre for the Study of Existential Risk, University of Cambridge

Abstract

In recent papers, Elizabeth Finneron-Burns and Johann Frick have both argued that it is not a wrong-making feature of human extinction that it would cause many potential people with lives worth living never to be born, and hence that causing human extinction would be, in at least one way, less wrong than many have thought. In making these arguments, both assume that merely possible future people cannot be harmed by their nonexistence, and thus do not have any claim to be brought into existence. In this paper, we raise objections to their arguments and suggest that there is nothing inherent in the moral theories they put forward that implies future people cannot have this sort of ‘existential’ claim. In doing so, we draw on the work of Derek Parfit, who argued, in a recent paper, that coming into existence benefits a person in that it is ‘good for’ them, even if it is not ‘better for’ them than non-existence. We also find that many of their objections to the view that it is wrong not to bring future people into existence rest on the assumption that, were these people to have claims on us, these must be equivalent to the claims that existing people have to be benefitted. However, we show that Parfit’s work demonstrates how this is not the case.

Keywords: Contractualism, Future generations, Human extinction, Non-identity problem, Person affecting principles.

“What I said was, ‘Is anybody at home?’” called out Pooh very loudly.

“No!” said a voice; and then added,

“You needn’t shout so loud. I heard you quite well the first time.”

“Bother!” said Pooh. “Isn’t there anybody here at all?”

(Winnie the Pooh)

1. Introduction

Most people agree that human extinction would be very bad and that causing, or failing to prevent it would, therefore, be very wrong. However, the reasons why it would be wrong, and the precise degree of its wrongness, remain a topic of debate amongst philosophers.

According to the most common argument put forward by those who believe that avoiding human extinction should be a global priority, the scale of human extinction’s badness lies in the number of potential future people who will not

come into existence if humanity goes extinct and who would otherwise live good and happy lives. This would make the end of humanity “the greatest of conceivable crimes” (Sidgwick 1907: 487). There is more than one approach to evaluating these future lives, but all agree that their value vastly outweighs the other negative effects of human extinction in both quantitative and qualitative terms (Leslie 1996; Bostrom 2003; Cirkovic 2004: 245, 252ff).

On one account, the value of these forgone lives lies in the well-being or utility they would contain (or more broadly in their Quality of Life). This view is supported by many consequentialists, including those who subscribe to Total Utilitarianism and Critical Level Utilitarianism but also by anyone who accepts what Derek Parfit terms ‘The Simple View’; that “Anyone’s existence is in itself good if this person’s life is worth living. Such goodness has non-diminishing value, so if there were more such people, the combined goodness of their existence would have no upper limit” (Parfit 2016: 112). On another account, the value of such lives is not that they contribute to the total quantity of any particular value, like utility, that our moral theory holds to be of primary importance, but that they will contribute to the quantity of any number of values that many different moral theories might appeal to. So long as we are even somewhat convinced to take any of these values into account in our moral deliberations, it is then argued, the sheer size of this effect should weigh heavily upon us whatever other moral beliefs we may hold. This can be termed its Expected Moral Value (Lockhart 2000; Greaves and Ord 2017).

However, some philosophers have objected to such arguments on the grounds that the existence, or non-existence, of merely possible future people should have no weight in our moral deliberations. Most notable among these dissenting voices is Thomas Michael Scanlon. According to his preferred brand of Contractualism, “an act is wrong if its performance under the circumstances would be disallowed by any set of principles for the general regulation of behaviour that no one could reasonably reject for informed, unforced, general agreement” (Scanlon 1998: 153). This is to say; a principle is morally sound just in case it is justifiable to each person, where such justification is attentive to the interests, complaints and other claims of the individuals in question.

Importantly, Scanlon places two restrictions on the reasons that can be offered for rejecting a moral principle. Firstly, there is:

The Impersonalist Restriction: impersonal values are not themselves grounds for reasonable rejection (Scanlon 1999: 222).

This means that if it is wrong to bring future people into existence then this must be because of the value of this existence to these people, it cannot simply be a recognition that these people’s lives would be good from some third-party point of view—as Sidgwick memorably put it, ‘the point of view of the universe’ (Sidgwick 1907: 382). Moreover, there is:

The Individualist Restriction: the justifiability of a moral principle depends only on individuals’ reasons for objecting to that principle and alternatives to it (Scanlon 1999: 229).

This means that the value of future people's lives to them must rest in the fact that this promotes some interest that these people actually have, and that we are failing to take into account in the outcome in which these individuals never exist.

In two recent papers, Elizabeth Finneron-Burns and Johann Frick have joined Scanlon in arguing that, the wrongness of human extinction cannot be a consequence of the fact that it will lead to many people never coming into existence, because people cannot have an interest in being brought into existence. To be sure, both then go on to advance arguments to the effect that causing, or failing to prevent, human extinction would nevertheless be wrong for other reasons. For instance, Finneron-Burns argues that human extinction is wrong because it will be bad for some people who already exist, for instance by leading to their premature death, while Johann Frick argues that its wrongness lies not in its impact on any individual human but what it will mean for humanity as a whole, thus preserving the individualist restriction whilst undermining the impersonalist restriction. However, our goal in this paper is not to engage with these positive suggestions (with which we broadly agree), but rather to interrogate the arguments that both authors offer for why we may not also need to consider the implications of our actions on merely possible people. In particular, we will argue that the view that there is some inherent feature of this kind of contractualist position which prevents those who accept it from giving any weight to the interests of merely possible future people is false. We thus conclude that if Finneron-Burns and Frick sincerely believe that the loss of potential future lives really is not a wrong-making feature of human extinction, then this will require some further kind of justificatory argument for this position and that to the extent that neither author provides such arguments, they are insufficient to support their conclusions.

2. The First Argument

Finneron-Burns states her principle argument against the need to take account of the impact of human extinction on merely possible future people as follows:

[W]e can only wrong someone who did, does or will actually exist because wronging involves failing to take a person's interests into account. When considering the permissibility of a principle allowing us not to create Person X, we cannot take X's interest in being created into account because X will not exist if we follow the principle. By considering the standpoint of a person in our deliberations, we consider the burdens they will have to bear as a result of the principle. In this case, there is no one who will bear any burdens since if the principle is followed (that is if we do not create X), X will not exist to bear any burdens. So, only people who do/will actually exist can bear the brunt of a principle, and therefore occupy a standpoint that is owed justification (Finneron-Burns 2017: 331).

According to this argument we can only wrong people by failing to take their interests into account (i.e. the individualist restriction), and people's interests only extend to not being burdened. Since future people do not have interests, as people who do not exist do not bear any burden by virtue of their non-existence, we, therefore, cannot wrong them.

There are several different replies that we might make to this argument.

On one of these, it is claimed that we can fail to take a person's interests into account, and hence wrong them, even in cases in which we do not, in fact, harm

or burden them in any way. One example in which we can be said to do this is when we act negligently, or even viciously, towards a person but fail to do any harm through sheer dumb luck. If this is possible, the argument goes, why is it not equally possible for us to be failing to take somebody's interests into account in cases in which that person is not burdened because they do not actually exist? (Kumar 2018).

In an earlier paper, Finneron-Burns objects to this line of argument on the grounds that while it is possible to wrong a person without harming them, this can only be achieved if there is a relevant 'standpoint', or cluster of personal characteristics and interests that an individual can occupy, whose interests are not being taken into account. However, she argues that there are a great variety of such standpoints that may be morally salient. Any of these could give us reasons for rejecting a moral principle on the grounds that it 'wrongs' a person who might occupy this standpoint, even though it does not actually harm anyone because nobody occupies it. However, she goes on to claim that this approach cannot be applied to the interests of merely possible people because standpoints can only describe positions that could be occupied by actual people and existence is not an interest that can attach to a standpoint (Finneron-Burns 2016).

According to another reply, it is claimed that we fail to take individual's interest into account in cases when they do not exist because non-existence is equivalent in value to a person existing with a neutral life, i.e. a life that is neither good nor bad for them, so that they would be genuinely better off if they had a good life instead. On this account, we can be said to burden merely possible future people because we have brought about a state, non-existence, that is worse for them than a possible happy life, which they might have had if we had acted otherwise (Fleurbaey and Voorhoeve 2015; Arrhenius and Rabinowicz 2015). This makes non-existence a real burden compared with the possibility of existence with a good life, and if we chose that a person should suffer this burden, then we harm them in the same way as if we had simply made their lives much worse so that they were no longer good for them.

However, from a Scanlonian perspective this raises the troubling prospect of ethical paralysis. Since everybody who does not come into existence with a life worth living is, on this account, harmed in just the same way as a person forced to live a neutral life, then it seems that all non-existing people have strong reasons to reject any principle that does not bring them into existence. Since all, or virtually all, principles may have some effect upon the identity of those who exist in the future, this would mean that there is always someone whose interests we are failing to take account of when applying such principles, and hence no, or almost no, ethical principles will be left on the table. As Finneron-Burns puts it "even basic, everyday activities would be morally wrong since they prevented certain people from coming to exist" (Finneron-Burns 2016: 1157).

We believe that there is another, better, reply to this argument that explains the wrongness of human extinction based on how it would harm future generations, but in a different way to other kinds of harm. According to this response, we harm, and therefore wrong, a person whom we do not bring into existence, not because we bring about a state of affairs that has zero value for them, but because we do not bring about any state of affairs that would have any, positive, value for them when we could have done. When we make choices that determine whether a possible future person will come into existence with a worth living life, we chose between an outcome that will be good for them and one that will not be

good for them—even though, since they do not exist in this second outcome, it will not be worse for them either. To put it another way, on this view we can be said to fail to take a person's interests into account, and hence to wrong them, because we have failed to benefit them, even though we have not burdened them either. This kind of argument was first proposed by Krister Bykvist (Bykvist 2007). However, its moral force has recently been more fully explored in a posthumously published paper by Derek Parfit, which will be our primary focus in this paper (Parfit 2017).

3. Parfit's View

Derek Parfit's opinion about arguments such as the one put forward by Finneron-Burns, which he terms the 'No Complainants Claim' is not easy to discern. He offers only one brief discussion of the position, summarised by the slogan "(a)n act cannot be wrong unless there is or will be someone whom this act has wronged" (Parfit 2017: 136). He admits that the view has both considerable psychological and moral force. However, he is keen to point out that it only describes part of morality. In particular, it rests upon the idea that a person's claims on us stem from the reasons they give us not to act badly towards them, what Parfit refers to as 'non-malevolence'. In Scanlonian terms, these are the reasons a person might have for rejecting the principle on which we are acting. In the case of existence, this is problematic, because we might say that a non-existent person cannot be a giver of reasons for anything.

In general Parfit believed that the differences between Scanlonian contractualism and consequentialism were not deep; however, in earlier writings, he appears to accept this line of argument and concluded that the wrongness of failing to create future people cannot stem from reasons that these people themselves would give us to bring them into existence. For instance, he argued that "Since [because of the non-identity problem] we cannot appeal to the personal reasons that are had by people who never exist, we should appeal to the impartial reasons that are had by people who do exist" (Parfit 2011, Vol. 2: 240).

However, in his final paper, Parfit moved beyond this position and argued that we could have personal reasons to bring people into existence. These would be reasons of benevolence to do what is good for a person and promote their personal goodness, and the difference between these two kinds of reason seems, at most, a matter of degree rather than of type, so that if a moral theory holds that we should respect reasons of non-malevolence it should not also claim that reasons of benevolence have no moral weight (Parfit 2017: 138).

Yet, following these rather inconclusive remarks, Parfit simply announces that he will not discuss this view further, because it is intrinsically related to an action's 'permissibility' which is not his topic of discussion at this point. He says that he will "return to the no complainants claim", but, unfortunately, he never does. This leaves Parfit's ultimate view about the wrongness of failing to bring a person into existence incomplete.

It is not our intention in this paper to attempt to put words into Parfit's mouth. However, in the absence of any conclusive statement of his views about this no complainants claim, we wish to explore a line of argument that these passages of his work suggest.

Parfit says that we should accept a 'Wide-Person Affecting Principle' according to which:

One of two outcomes would be in one way worse if this outcome would be less good for people, by benefiting people less than the other outcome would have benefited people (Parfit 2017: 129).

Unlike the No-Complainants Claim, this principle avoids limiting us to reasons for non-malevolence but considers the whole of morality, including how we might benefit people. Of course, there is no great difficulty in saying that a person whom we might bring into existence and who would have a good life could give us reasons to choose to bring them into existence, on the grounds of benevolence, because they would be glad to be alive in the same way that a person whom we might bring into existence who would have a bad life could give us reason to choose not to bring them into existence, on the grounds of non-malevolence, because they would regret our actions.

Finneron-Burns claims that ‘wronging involves failing to take someone’s interests into account’, however, in arguing that we cannot wrong future generations she equates a person’s interests with bearing ‘burdens’. Yet, many would say that avoiding burdens constitutes, at most, half of a person’s interests and that our interests can equally be served by receiving ‘benefits’. When we fail to bring somebody into existence, therefore, we do not have to say that we have failed to take their interests into account because, in the outcome in which they do not exist, their interests are not being ignored; rather we can say that we are ignoring the interests of the person in the alternative outcome in which they do exist, with a good life, and are glad that they do. If we then adopted a principle of action that would mean that these people no longer exist, then it is these reasons that count in favour of rejecting that principle, not the supposed reasons given by any ‘non-existent’ people as Finneron-Burns suggests. Since one of the outcomes that we might choose is good for these people, and hence serves their interests and benefits them, failing to bring this outcome about would be bad for these people, fails to take account of their interests and thus constitute a kind of harm for them, even though it is clearly not ‘worse’ for them. Hence, we can be said to have done wrong by failing to act in the way that would have been good for these people, and thus denied them this existential benefit, whether or not the interests we fail to take into account can be said to actually exist in the outcome that we eventually bring about.

Of course, Finneron-Burns and Scanlon have their reasons for preferring to focus on moral burdens. One of these is surely that their moral philosophy, in general, has a negative tone, since the aim of Scanlonian contractualism is to identify which principles are morally ‘impermissible’ by working out which ones people could ‘reasonably reject’. Parfit himself has done much to critique this view, and to argue that it must be counterbalanced by a moral philosophy that considers which principles are required by virtue of being the ones that people could ‘rationally accept’, although ultimately, he also believes that the two approaches to morality arrive at the same result—or at least that they are climbing the same mountain (Parfit 2011). However, we do not think that the point we are concerned with here requires us to accept this further claim of Parfit’s. Perhaps the theoretical equivalence between impermissibility, reasonable rejection and burdening is of symbolic importance. Yet, there is no reason why it would not be of equal moral importance to equate impermissibility, reasonable rejection and not benefiting. Indeed, for the kinds of ‘same population’ cases with which Scanlonian contractualists are overwhelmingly concerned, the two would appear

to amount to very much the same thing, and it is only when one considers ‘different population’ cases, and in particular the interaction between Scanlonian contractualism and Wide Person-Affecting views, that it becomes morally salient. To then argue that in these cases one can only go along with the original concept of moral burdens, and claim that since nobody is burdened by their non-existence, then causing them not to exist cannot be wrong, appears to us to be going too far. We would have much less of a difficulty in saying that there are people who are not benefited because they do not exist: why then should we not owe some justification to these people for failing to do what would have been good for them, and why would that not provide an equal ground for wrongness via reasonable rejection?

4. Other Arguments

However, this is not Finneron-Burns’ only argument against the wrongness of failing to bring people into existence. Her second argument runs as follows:

[E]xistence is not an interest at all and a possible person is not disadvantaged by not being caused to exist. Rather than being an interest, it is a necessary requirement in order to have interests. Rivka Weinberg describes it as ‘neutral’ because causing a person to exist is to create a subject who can have interests; existence is not an interest itself. In order to be disadvantaged, there must be some detrimental effect on your interests. However, without existence, a person does not have any interests so they cannot be disadvantaged by being kept out of existence. But, as Weinberg points out, “never having interests itself could not be contrary to people’s interests since without interest bearers, there can be no ‘they’ for it to be bad for” (Finneron-Burns 2017: 331).

This argument appears at first reading to equate the claim that we might wrong a person by failing to bring them into existence with the claim that people have some kind of intrinsic interest in their existence. However, this is clearly not what those who are concerned about the wrongness of failing to bring people into existence are arguing for. The reasons for this are not hard to see if we consider the possibility of bringing somebody into existence whose life would be bad for them, and worse than nothing. Such a person’s existence is clearly not in their interest; rather it would be in this person’s interests for us not to be brought into existence. It is not clear from Finneron-Burns’ argument whether she believes that we would wrong a person by bringing them into existence with such a bad life, although we suspect that she would. On her view, this person clearly would have been burdened by being brought into existence and would be present in the outcome under consideration to regret this outcome, to demand justification for it, and hence to be wronged by it on any account. It would follow from this therefore that the mere fact that a benefit or burden stems from being brought into existence should not, in and of itself, be seen as an objection to its potential wrongness. It is not, to borrow Finneron-Burns’ own words, that this person’s existence is not in their interests per se, but merely that, since the life they live would be bad, it would be less bad (or more good) for this person never to have existed.

The reason for this is surely that the interest a person has in their existence is not related to the existence itself, but to the kind of life that they will lead. Indeed, there is a kind of life, that at the so-called ‘neutral level’ that even the most ardent

supporter of the interests of merely possible people would claim is in nobody's interests to either have or to avoid.

This argument cannot then be interpreted as concerning whether people have an interest in their existence or not. Instead, we believe that the best way of interpreting this argument is as an objection to the notion that a person who does not exist in a particular outcome can have any moral standing regarding the value of that outcome, even relative to the value of other outcomes in which they would exist. However, the strength of this objection surely depends upon our notion of how and why individuals have a bearing on the value of outcomes. Finneron-Burns', and Scanlon's account both offer us a clear position on this question, since it seems that only those who can strictly be described as being burdened can have any such moral standing. As Finneron-Burns puts it "a principle that results in some possible people never becoming actual does not impose any costs on those 'people' because nobody is disadvantaged by not coming into existence" (Finneron-Burns 2017: 331). However, this is surely begging the question against the view that people can have an interest in what is good for them, whether or not it is better or worse for them than something else. If we adopted a slightly different approach to contractualism in which one can just as well impose a cost on somebody by failing to do what is good for them, then Finneron-Burns' second argument no longer follows. The strength of this objection seems to depend entirely upon the success of Finneron-Burns' first argument, rather than offering any fresh objection to the view that it can be wrong to perform an action that fails to benefit merely possible people.

In a footnote, Finneron-Burns offers a third objection to the view that we wrong future people by failing to bring them into existence. This is that "Depending on how much weight we assign to the 'good' of existence, to say that a possible person could reasonably object to not being created would make every action or principle wrong if it outweighed the burdens we would have by not creating 'them'" (Finneron-Burns 2017: 341). This is an important point. However, as Finneron-Burns notes, it depends entirely upon how we weigh this good. The difficulties that Scanlonian contractualism has in weighing up different goods and giving priority to individual claims are widespread and not confined to different population cases. However, even if one believed that these were unusually problematic for the contractualists position, it would seem ad hoc to exclude the claims of merely possible people purely on the basis that they were hard to aggregate.

Indeed, we believe that one of the key advantages of Parfit's means of avoiding the Non-Identity problem is that: by distinguishing how we can make an outcome 'worse' for people and how we can make it simply 'less good' for them, we avoid the need to equate the claims of actual and merely possible people as being of the same kind, and hence deserving identical moral weight. It is entirely open to us to say that, even though the benefit we might give to merely possible people is huge, it should not outweigh the lesser benefits that we might otherwise give to actually existing people, because these are two different kinds of benefits leading to two different kinds of claim, and that people's claims to 'comparative' benefits take (a greater or lesser) degree of priority over their claims to existential ones.

5. Frick's View

In another paper, presented in the same volume as Finneron-Burns', Johann Frick presents a different account of the morality of human extinction. Frick also rejects

the notion that extinction is bad because of the value of the good lives it would forgo, although he does suggest that these might make it wrong by reducing what he terms the 'final value of humanity'. As with Finneron-Burns' argument, however, it is not our intention to engage with these positive arguments (which we broadly agree with) but only with his rejection of the value of the lives of merely possible future people.

Like Finneron-Burns, Frick is a contractualist. However, this appears to be less significant to his reasons for rejecting the value of future lives. Indeed his primary argument involves developing and defending an intuitive belief he terms the 'intuition of neutrality'. Frick defends this intuition from an often cited objection to it, John Broome's 'Greediness Objection' (Broome 2004). However, he also offers some positive arguments in favour of this intuition, and by extension against the view that it would be wrong to create extra happy lives, which we will now consider.

The intuition of Neutrality, as stated by Broome, runs as follows:

[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).

Frick argues that we should accept this intuition for three reasons.

The first of these is that the intuition supposedly "neatly dovetails with a deontic intuition about the morality of procreation, the so-called Procreation Asymmetry" (Frick 2017: 348). According to this

[I]f 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 (Frick 2017: 348).

We say that these two intuitions only 'supposedly' dovetail because there are some important differences between them. The first of these is that on many versions of the intuition of neutrality future people with lives that are sufficiently good, i.e. above what is known as the 'neutral range', should be created if we have the option to do so. However, the procreative asymmetry would instead imply that there is still no reason to bring these people into existence. Frick acknowledges this and states that he personally prefers versions of the intuition of neutrality on which the neutral level has no upper bound of this kind. However, this does not remove the fact that there is an inconsistency between these views, with the intuition of neutrality still allowing that people may have obligations to create extra people with (very) happy lives, while the procreative asymmetry allows for no such thing.

Another point at which the two fail to dovetail is that our reasons for accepting the intuition of neutrality stem from one set of considerations, the lack of intrinsic value in the lives of merely possible future people. However, we think that for many people the procreative asymmetry at least partially reflects another a different consideration, the interest that a prospective parent has in deciding for themselves whether to have children, free from most moral considerations. In particular, many people feel that there are no moral conditions under which individuals who do not want to have children should be required to have them. For

instance, the intuition of neutrality is compatible with the belief that if one does decide to have a child, then one ought to have the happiest child that one might have, since in this case, we are not assessing the value of adding an extra person, but of adding one particular extra person or another. However, according to the procreative asymmetry, the degree to which a merely possible future person's life is worth living presents us with no reason to create that person, rather than creating any other person (so long as their life is not 'not worth living') as well as creating no person at all.

The second reason that Frick gives in support of the intuition of neutrality is that rejecting it appears to have 'strongly revisionary implications' of its own. For this, Frick gives another example from John Broome:

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

This intuition appears to dovetail even less well with the intuition of neutrality (and the contractualist position) than the procreation asymmetry. This is because, as should be fairly obvious, when people choose not to consider the effects of our actions on the lives of potential future people they do not distinguish between whether these are 'good lives' with positive welfare or 'bad lives' with negative welfare. However, the intuition of neutrality only says that we should be neutral about bringing into existence people with good lives (and on many interpretations only those with lives within a neutral range with an upper bound). If we in fact used the intuition of neutrality in the case that Broome describes we should acknowledge that at least some small fraction of the future population that we might create by saving people's lives through road safety measures could potentially have bad lives, and this badness *should* have an impact on our evaluation of the scheme as a whole, even if the goodness of the other lives should not. This would imply a version of what Derek Parfit terms his 'absurd conclusion': that even when we expect a new population to have broadly the same characteristics as an existing population, which we take to be good, the addition of these new lives will make an outcome containing them worse, because the impact of the (small number of) negative lives on our evaluation of this outcome will outweigh the impact of the (great number of) good lives (Parfit 1984).

We think that there may be a more compelling reason for why people do not consider the population effects of interventions like road safety when we evaluate them, namely the fact that we have so little information about these lives and their overall value as to render us 'clueless' about how to evaluate them (Greaves 2016). If indeed we ignored this wellbeing because we accepted the intuition of neutrality then we should ignore it even when we are considering measures that are intended to improve the quality of life for those yet unborn. For instance, we should not consider the value of genetic research or screening that will save future people from suffering disabilities, except in so far as these disabilities impact on the quality of life of their parents. Yet in such cases, where the aim and purpose of our actions is to improve the lives of future people so that we are no longer

clueless about our expected effect upon them, we would routinely consider our effect upon their lives in any cost-benefit analysis (although, tellingly, we may still not consider our effect upon their descendants, since we are still clueless about this).

Frick's final reason for accepting the intuition of neutrality is that it helps to avoid some of the objectionable implications of population axiology, most notably Derek Parfit's Repugnant Conclusion. While this is true, it is a general fact of theories of population ethics that they all have some such objectionable implications (Parfit 1984; Arrhenius Forthcoming). For instance, as we have seen, the standard form of the intuition of neutrality implies versions of the Absurd Conclusion, whilst a fuller version of the intuition of neutrality, which held that no future lives had any value, would imply what Parfit terms the Callous Conclusion—that a future population of people whom all have bad lives and suffer a great deal would be no worse than a population of people whom all have good lives.¹

Furthermore, and as stated above, merely allowing that future people's lives have value, whether one is a contractualist or a consequentialist, does not have to imply anything about how to weigh this against other valuable things or these claims against other claims. In extending his Wide-Person Affecting Principle, Parfit developed one approach that demonstrates this well, which he labelled his Wide Dual Person-Affecting Principle:

One of two outcomes would be in one way better if this outcome would together benefit people more, and in another way better if this outcome would benefit each person more (Parfit 2017: 154).

Parfit argued that this principle would give better results when applied to population ethics type cases, and yet it is still compatible with the contractualist account of wrongness and the claims of potential people. Ultimately however, Parfit did not believe that this principle could necessarily escape the Repugnant Conclusion on its own, although he believed that, with other principles, it could form part of a moral theory that would—his Theory X. We believe that such a theory is worth exploring and that, as Parfit clearly believed, both consequentialists and contractualists might have something to add to this effort.

6. Conclusions

Ultimately, we do not believe that any of the arguments and points we have made here could force a contractualist, such as Finneron-Burns or Frick, who did not want to adopt the view that human extinction was wrong, at least in part, because it meant forgoing many potential future lives to change their mind. However, nor do we believe that the arguments we present here are the only ones that could achieve this end (we present further arguments, grounded in other obligations, in Kaczmarek and Beard forthcoming). We do hope that we have raised some doubts as to whether this is the position that a contractualist is obliged to take regarding the evaluation of its wrongness. We believe that for a contractualist to

¹ The name 'callous conclusion' comes from an unpublished, although widely circulated, paper titled 'Towards Theory X'. However, the intuition that this would be a bad conclusion is widely shared (see e.g. Broome 2005 and Arrhenius forthcoming).

say that potential future people have no claims against us is a legitimate position to take and that it is the analogue of consequentialist positions like ‘Presentism’ or ‘Actualism’, which say the same thing. However, we think that it is merely one of many legitimate positions that a contractualist might take on the matter. Many people think that the way in which these theories evaluate human extinction is one of the bad features of these theories, and that it gives us reason to prefer theories that give more weight to the value of future lives (indeed this is the motivation for Finneron-Burns’ and Frick’s positive proposals regarding the badness of extinction). Perhaps Contractualists might eventually come to say the same things about versions of contractualism according to which merely possible future people have no claims. Alternatively, perhaps they will not.²

References

- Arrhenius, G. (forthcoming), *Population Ethics: A Challenge for Moral Philosophy*.
- Arrhenius, G., & Rabinowicz, W. 2015, “The Value of Existence”, in Hirose, I. and Olson, J. (eds.), *The Oxford Handbook of Value Theory*, Oxford: Oxford University Press, 424-43.
- Bostrom, N. 2003, “Astronomical Waste: The Opportunity Cost of Delayed Technological Development”, *Utilitas*, 15, 3, 308-14.
- Bostrom, N. 2013, “Existential Risk Prevention as Global Priority”, *Global Policy*, 4, 1, 15-31.
- Broome, J. 2004, *Weighing Lives*, Oxford: Oxford University Press.
- Broome, J. 2005, “Should We Value Population?”, *Journal of Political Philosophy*, 13, 4, 399-413.
- Bykvist, K. 2007, “The Benefits of Coming into Existence”, *Philosophical Studies*, 135, 3, 335-62.
- Ćirković, M.M. 2004, “Forecast for the Next Eon: Applied Cosmology and the Long-term Fate of Intelligent Beings”, *Foundations of Physics*, 34, 2, 239-61.
- Finneron-Burns, E. 2016, “Contractualism and the Non-identity Problem”, *Ethical Theory and Moral Practice*, 19, 5, 1151-1163.
- Finneron-Burns, E. 2017, “What’s Wrong with Human Extinction?”, *Canadian Journal of Philosophy*, 47, 2-3, 327-43.
- Fleurbaey, M. & Voorhoeve, A. 2015, “On the Personal and Social Value of Existence”, in Hirose, I. and Reisner, A. (eds.), *Weighing and Reasoning: Themes from the Philosophy of John Broome*, Oxford: Oxford University Press. DOI: 10.1093/acprof:oso/9780199684908.001.0001
- Frick, J. 2017, “On the Survival of Humanity”, *Canadian Journal of Philosophy*, 47, 2-3, 344-67.

² We are grateful for discussions with Elizabeth Finneron-Burns, Johann Frick, Michael Plant and Derek Parfit that helped to inform this paper. We are also grateful to an anonymous reviewer for this journal for their feedback and to Phil Torres for help with proof-reading. This project was made possible through the support of grants from Templeton World Charity Foundation and the Berkley Existential Risk Initiative. The opinions expressed in this publication are those of the author(s) and do not necessarily reflect the views of these foundations.

- Kaczmarek, P. & Beard, S. (forthcoming), "Human Extinction & Our Obligations to the Past", *Utilitas*, 1-10, DOI: 10.1017/S0953820819000451
- Kumar, R. 2018, "Risking Future Generations", *Ethical Theory and Moral Practice*, 21, 2, 245-57.
- Greaves, H. 2016, "Cluelessness", *Proceedings of the Aristotelian Society*, 116, 3, 311-39.
- Greaves, H. & Ord, T. 2017, "Moral Uncertainty about Population Axiology", *Journal of Ethics and Social Philosophy*, 12, 2, 135-67.
- Leslie, J. 1996, *The End of the World: The Ethics and Science of Human Extinction*, New York: Routledge.
- Lockhart, T. 2000, *Moral Uncertainty and Its Consequences*, Oxford: Oxford University Press.
- Parfit, D. 1984, *Reasons and Persons*, Oxford: Oxford University Press.
- Parfit, D. 2011, *On What Matters, Volume 1 and 2*, Oxford: Oxford University Press.
- Parfit, D. 2016, "Can We Avoid the Repugnant Conclusion?", *Theoria*, 82, 2, 110-27.
- Parfit, D. 2017, "Future People, the Non-Identity Problem, and Person-affecting Principles", *Philosophy & Public Affairs*, 45, 2, 118-57.
- Sidgwick, H. 1907, *Methods of Ethics*, London: Macmillan & Co.
- Weinberg, R. 2008, "Identifying and Dissolving the Non-identity Problem", *Philosophical Studies*, 137, 1, 3-18.