

The Heart of the Problem with Longtermism

Ben Bramble
Australian National University
ben.bramble@anu.edu.au

Draft as of 1st September, 2023

1 Introduction

William MacAskill defines longtermism as the view that “positively influencing the longterm future is a key moral priority of our time”¹. This definition makes longtermism sound modest, even platitudinous. What reasonable person *doesn't* think we should be, say, urgently addressing threats to the planet's longterm health or survival, like climate change, nuclear war, stray asteroids, and so on?

But make no mistake, longtermism is not a modest proposal—it is radical indeed.² This is because of what longtermists have in mind by the ‘longterm’. While many of us who care about the future are thinking roughly of the next thousand years or so (if we have a timeframe in mind at all), longtermists are thinking—explicitly—of the next *trillion* years or so.

Defenses of longtermism usually start with the observation that we stand today at what is potentially only the very beginning of human history. If humanity keeps on going, we could spread throughout the universe, or even upload ourselves into virtual universes, and survive for trillions of years. In this time, there could be trillions upon trillions more humans (or intelligent

¹ William MacAskill, *What We Owe the Future* (Basic Books, 2022), 12.

² Longtermists prefer “revolutionary” (MacAskill, *ibid.*, 16).

beings descended from humans), on some estimates as many as 10^{54} .³ With suitable advances in technology, these future humans could have lives vastly higher in well-being than any of us today.

What is the relevance of this observation, for longtermists? Why is it so important? You could be forgiven for thinking that longtermists' concern is *to make things better for these many future humans*. They often speak, after all, of 'the moral irrelevance of temporal distance'. Here is MacAskill:

Future people, after all, are people. They will exist. They will have hopes and joys and pains and regrets, just like the rest of us. They just don't exist *yet*... People matter even if they live thousands of miles away. Likewise, they matter even if they live thousands of years hence.⁴

Similarly, Toby Ord writes:

People matter equally regardless of their temporal location... Our lives matter just as much as those lived thousands of years ago, or those a thousand years hence. Just as it would be wrong to think that other people matter less the further they are from you in space, so it is to think they matter less the further away from you they are in time. The value of their happiness, and the horror of their suffering, is undiminished.⁵

Longtermists also refer in this context to Peter Singer's notion of "the expanding circle", the idea that moral progress consists, at

³ Nick Bostrom, "Existential Risks FAQ" (2013), accessible here: <https://existential-risk.org/faq.pdf>.

⁴ MacAskill, *ibid.*, 17.

⁵ Toby Ord, *The Precipice* (Bloomsbury, 2020), 52. Hilary Greaves, also, in motivating longtermism, says: "If there's a child suffering terribly in 300 years' time, and this is completely predictable—and there's just as much that you could do about it as there is that you could do about the suffering of a child today—it'd be pretty strange to think that just because it's in the future it's less important." (CEPPA Chats) Accessible here: <https://www.youtube.com/watch?v=d1jMlb8E08k>.

least in part, in the expansion of altruism “from the family and tribe to the nation and race”, and then to all other humans regardless of ethnicity, gender, sexual preferences, and so on, and then out still further to non-human animals.⁶ According to longtermists, the next big step in this process of expansion is to include *future* beings. As Ord puts it, “Recognizing that people matter equally, wherever they are in time, is a crucial next step in the ongoing story of humanity’s moral progress.”⁷

Despite such talk, however, the longtermist’s concern is not—or, at least, not fundamentally—with making things better for the future trillions. To see this, suppose we learned that a massive asteroid is on course to hit Earth in 200 years from now. If we do nothing, it will obliterate the planet. Would longtermists think “Since there now *won’t* be trillions upon trillions of future humans, we needn’t worry about spending money on the future”? No. Instead, they would say that we should be urgently spending money on averting this catastrophe, so that humanity can carry on into the future.

The longtermist’s fundamental concern, you see, is not with making things better for the future trillions, but with making it more likely *that there will be such trillions in the first place (assuming they will be happy)*. Their overarching aim, put crudely, is to *further populate the distant future*. MacAskill makes this clear in a section of his book labelled ‘Bigger Is Better’, where he writes:

We should...hope that future civilisation will be *big*. If future people will be sufficiently well-off, then a civilisation that is twice as long or twice as large is twice as good...The future of civilisation could be literally astronomical in scale, and if we will achieve a thriving, flourishing society, then it would be of enormous importance to make it so.⁸

The reason longtermists are so concerned with the distant future is not out of a concern for future people *per se*, but from a

⁶ Peter Singer, *The Expanding Circle* (Clarendon, 1981), 120.

⁷ Ord, *ibid.*, 52.

⁸ MacAskill, *ibid.*, 201.

realisation of just how much happiness—and, so, by longtermists’ lights, how much *value*—these vast future years could potentially contain.

At the heart of longtermism is the following idea, familiar from population ethics, which I will refer to as *the Heart*:

There is intrinsic value in the addition of each new happy being to the world.

The idea is that adding more happy beings to the world makes the world better *in and of itself*—i.e., independently of any effects or other implications doing so might have.⁹

Longtermism exists today because a bunch of Oxford philosophers who already accepted the Heart (or had a sufficiently high credence in it) came to recognise that the longterm future could contain trillions upon trillions of new happy beings, and recognising this, decided to put their money where their mouth is and advocate for (what they took to be) an implication of these ideas.

It is important to note that many longtermists do not think that adding more happiness to the world is the *only* thing that matters. MacAskill, for example, emphasises that he now “[rejects] the utilitarian [idea that] the ends always justify the means”, and endorses “moral side-constraints (don’t violate people’s rights!)”¹⁰. He also believes in the existence of special reasons of *partiality* and *reciprocity*, which, he says, entail that we should give extra weight to the interests of current people (given that we

⁹ MacAskill puts it like this: “Provided a person had a sufficiently good life, the world would be a better place in virtue of that person being born and living that life. Crucially, this isn’t the claim that an additional person might make the world better by enriching the lives of others; instead, it’s the claim that having one extra person in the world is good in and of itself, if that person is sufficiently happy...All other things being equal, having more happy people makes the world a better place.” (MacAskill, *ibid.*, 177.)

¹⁰ Tweeted August 16, 2022. Accessible here: <https://twitter.com/willmacaskill/status/1559196018062786560>.

are partial and indebted to them).¹¹ Ord and other longtermists hold similar views.¹²

Note, also, that most longtermists do not, right now, advocate policies so very different from those most people today already endorse—namely, addressing threats like climate change, nuclear war, stray asteroids, and so on, while at the same time improving current institutions and living conditions for people alive today. The reason for this alignment is that most longtermists accept that our ability to predict the very longterm effects of different policies is currently still quite poor. Consequently, they think, the safest bet at the moment for further populating the distant future is roughly these standard sorts of policies.¹³

Why, then, is longtermism so radical? It is because of *hypotheticals*. If it *were* clear what actions would not only address existential threats, but *also* further populate the distant future, then, on longtermism, we should spend or sacrifice huge amounts to take these actions. These sacrifices might not involve directly harming anyone—in line with MacAskill’s prohibition on such harming—but would nonetheless involve *failing to help* people today who are living with subpar healthcare, education, food, and so on.

You might object to longtermism on just these grounds. ‘Intuitively’, you might say, ‘if we were in this improved epistemic situation, we should *not* make such sacrifices to further populate the distant future. Such sacrifices, while they might not directly harm anyone, are nonetheless too great.’

¹¹ MacAskill, *ibid.*, 18.

¹² Ord, for example, writes: “We may have special duties to some individuals—to family; to members of the same community.” (Ord, *ibid.*, 52.)

¹³ MacAskill, for example, says: “Moving off carbon is a win-win for both the near and the long term. The same holds for preventing pandemics, controlling artificial intelligence and decreasing the risk of nuclear war.” MacAskill, “The Case for Longtermism”, *The New York Times*, August 5, 2022.

Accessible

here:

<https://www.nytimes.com/2022/08/05/opinion/the-case-for-longtermism.html>.

Unfortunately, this is not much of an argument. Longtermists will simply respond that while it might be counterintuitive to think we should make such sacrifices, longtermism has a solid theoretical basis, and we should follow good arguments where they lead.

To shift a longtermist, you might try to argue that any reasons of partiality or reciprocity we have in this context are *so* strong that even in an improved epistemic situation we should not make such sacrifices to further populate the distant future.

Alternatively, you might try to argue that, while there is indeed immense value in further populating the distant future, there is even more value to be achieved in helping existing people meet their needs (*however* many new future people we could add). That is, you might try to show that the value of existing people's welfare has what philosophers call *lexical* priority or superiority.

I will not pursue these approaches. I doubt very much there are any special reasons of partiality or reciprocity, and appeals to lexical priority seem in this context *ad hoc* (or at least, not very explanatory).

I will take a different approach. I will aim for the Heart. I will argue that longtermists are mistaken in the first place that there is intrinsic value in the creation of new happy beings. For this reason, there is not, after all, so much at stake in whether the distant future will contain trillions of happy people or not. On the contrary, there is *nothing* at stake in this *per se*. We cannot do *any* good, let alone huge amounts, by further populating the distant future.

I am not the first to object to longtermism on these sort of grounds.¹⁴ But I aim to go deeper into the literature on population ethics than others have been able to do in their contributions to popular media on the topic. In doing so, I will be offering a number of new contributions to this literature.

¹⁴ See, for example, Kieran Setiya, "The New Moral Mathematics" (Boston Review, August 2022), accessible here: <https://www.bostonreview.net/articles/the-new-moral-mathematics/>. See also Émile P. Torres, "Against Longtermism", *Aeon Magazine*, accessible here: <https://aeon.co/essays/why-longtermism-is-the-worlds-most-dangerous-secular-credo>.

There is an obvious reason, from the outset, to be wary of the Heart. This is that creating a new happy being seems not to *benefit* this being. Why does it not benefit them? It is because it does not make them any better off than they would otherwise have been. If they were not created, they would never exist at all, and the non-existent are neither well nor badly off in any ways. To many of us, there seems nothing of value in actions that benefit no one.¹⁵

Longtermists, however, are unimpressed by this point. They hold that creating new happy beings is intrinsically good *even if it doesn't benefit these beings*. What's good here, they say, is just that there will be more happiness or welfare in the world.¹⁶ To do good, it isn't necessary to benefit anyone—it is enough to add more welfare to the world. Let us call this the *impersonalist* version or conception of the Heart. Longtermists offer two main arguments for the Heart so conceived, each of which they take to be extremely strong: (1) the Argument From Extinction and (2) the Argument From Miserable Beings.¹⁷ It is their high level of confidence in the soundness of these arguments that ultimately explains the origin and persistence of longtermism as an ideology and movement.

The first half of this paper (Sections 2 & 3) will be devoted to offering new responses to these two arguments. I will then (Section 4) describe an alternative to longtermism, a view I call

¹⁵ See, for example, Jan Narveson, “Moral Problems of Population”, *The Monist* 57 (1973) (1): 62–86, and John Broome, *Ethics Out of Economics* (Cambridge University Press, 1999).

¹⁶ Some longtermists do think there is a sense in which creating a new being can be good for them even if it cannot be better for them. It can be good for them *noncomparatively*. This point was originally made by Derek Parfit in *Reasons and Persons* (Oxford, 1984), 488. Most longtermists, however, seem to follow Jeff McMahan in holding that this form of benefiting does not itself provide a reason to create new happy beings. Your reason to create new happy beings is not to benefit these beings, but to add more welfare to the world. See McMahan, “Asymmetries in the Morality of Causing People to Exist”, in *Harming Future Persons*, eds., Melinda A. Roberts and David T. Wasserman (Spring, 2009), 52.

¹⁷ There are several other arguments, too, but these two are the main ones. I address the others in separate work.

future sentimentalism, which seems to me to do a better job of explaining our future-regarding reasons. I will then (Section 5) consider an important objection to my critique of longtermism. Finally (Section 6), I will sum up my argument.

2 The Argument From Extinction

The longtermist's first argument for the Heart begins with a relatively uncontroversial idea, that human extinction would be an extremely bad thing, one of the worst things imaginable.¹⁸ Why would it be so bad? Part of the reason is that it would greatly harm the final generation. But this, according to longtermists, cannot fully explain its badness. Longtermists here cite Derek Parfit's *Two Wars*.¹⁹ Parfit asks us to consider three scenarios: (1) peace, (2) a war that kills 99% of the world population, and (3) a war that kills 100% (resulting in human extinction). According to Parfit (and now longtermists along with him), while (2) is obviously much worse than (1), the difference in badness between (2) and (3) is greater still. To fully account for the badness of (3), it is not enough to appeal to the additional harms caused to the final 1% of humanity. There must be a further factor.

Longtermists also point to cases like Larry Temkin's *Extinction Pill*:

If we developed a pill enabling each of us to live wonderful lives for 120 years, it would be terrible for us to take the pill if the cost of doing so were the extinction of humanity.

Moreover, this is so even if taking the pill were better for each individual who took it, and hence, collectively, for everyone who was alive then or later lived. We think the outcome where people lived wonderful lives for 120 years would be much

¹⁸ I say 'relatively' because of David Benatar, *Better Never to Have Been* (Oxford, 2006) and Émile P. Torres, *Human Extinction: A History of the Science and Ethics of Annihilation* (Routledge, forthcoming).

¹⁹ Parfit, *ibid.*, 452.

worse than the outcome where people lived lives of 80 years, but human life continued on for countless centuries.²⁰

In cases like Temkin's, longtermists say, extinction still seems bad, even though it *doesn't* harm the final generation.

Why are these extinctions so bad, according to longtermists? Intuitively, they say, it is because *they prevent the existence of all the happy people who would have lived had humanity kept on going*. Ord, for example, invites us to imagine, if our own generation were to be the last, the many

children and grandchildren we would never have: millions of generations of humanity, each comprised of billions of people, with lives of a quality far surpassing our own. *Gone*. A catastrophe would not *kill* these people, but it would foreclose their very existence. It would not *erase* them, but ensure they were never even written.²¹

This, Ord says, would be an immense loss.

Similarly, Jeff McMahan writes:

To most of us, it is appalling to think that instead of this incalculable number of people enjoying these incalculable benefits, there might instead be only the emptiness of a world devoid of consciousness.²²

And here are Nick Beckstead, Peter Singer, and Matt Wage:

One very bad thing about human extinction would be that billions of people would likely die painful deaths. But in our view, this is, by far, not the worst thing about human extinction. The worst thing about human extinction is that *there would be no future generations*...If we fail to prevent our

²⁰ Larry Temkin, *Rethinking the Good* (Oxford, 2012), 414. See also Jonathan Glover, *Causing Death and Saving Lives* (Penguin, 1977).

²¹ Ord, *ibid.*, 143.

²² Jeff McMahan, "Causing People to Exist and Saving People's Lives", *Journal of Ethics* (2013), 26.

*extinction, we will have blown the opportunity to create something truly wonderful: an astronomically large number of generations of human beings living rich and fulfilling lives.*²³

Longtermists conclude that there must be intrinsic value in the addition of each new happy life, and thereby arrive at the Heart.

I believe this argument for the Heart fails. To see why, consider:

Utopian Extinction. Many millenia into the future, after years of slow and steady progress, humanity finally achieves a glorious utopia, where everyone lives long lives as good as it is possible to live. At this point, there is nothing further for humanity to discover or accomplish in the arts, sciences, philosophy, morality, or in space exploration. Humanity really has done it all, and these humans all *have* it all. As such, they become, not bored exactly, but satiated, thoroughly content. They decide, collectively and unanimously, that they do not want to have children or raise a further generation. They would prefer to wrap up the human experiment, and go out on a high. One sunny day, they do so and humanity ends.

This extinction does not seem regrettable at all. Perhaps it is regrettable that ‘this is all there is’, cosmically speaking—that there is no bigger purpose, plan, or meaning for humanity, no afterlife that might make further sense of our lives, reunite us with past people, or bring us all closer together in some deep or significant way. But *given* that there isn’t anything bigger like this, I see nothing to regret in these utopians wrapping things up. However,—and here is the crucial point—this extinction also (like the Parfit/Temkin extinctions) *prevents the existence of trillions of new happy beings*. Given this, it cannot be that the loss of trillions of new happy beings *itself* makes an extinction

²³ Nick Beckstead, Peter Singer, and Matt Wage, “Preventing Human Extinction”, *Effective Altruism Forum* (2013), accessible here: <https://forum.effectivealtruism.org/posts/tXoE6wrEQv7GoDivb/preventing-human-extinction>.

bad. So, this loss cannot be something that is bad in the case of the Parfit/Temkin extinctions.

As further evidence, consider a second case:

Happiness Wand. You have a wand which, whenever it is waved, creates trillions of new happy beings *now* in a physically distant part of the universe, a part so remote no one will ever be able to visit or communicate with these beings at any time in the future. Only you will know they exist.

Suppose you choose not to wave this wand. Was this a bad thing to do? If there is intrinsic value in the addition of new happy beings, we should regard your failure as *extraordinarily* bad. But intuitively, it is not. It does not seem bad at all. You do not do something bad if you put down this wand and get back to your normal life.

So, what *is* going on in the Parfit/Temkin cases? What *are* we reacting to here? Longtermists, after all, do seem right that *something* deeply concerns us when we attend to such cases.

I believe that what we are reacting to here is the fact that these extinctions are extraordinarily *sad*. Our awareness of their sadness is leading some of us—including longtermists—to think, mistakenly, that they are *bad* over and above the harms they involve for the final generations. When something is sad, there is a tendency to slip into thinking of it as bad, or as making things worse *simpliciter*. But this is a slip. While *most* things that are sad are also bad, not *all* are. Crucially, *things can be sad without being bad*. This, I think, helps to explain what is going on in the Parfit/Temkin cases.

To better understand what I am suggesting, consider another sort of case where something is sad without being bad. Suppose an old person with no living friends or relatives passes away, and her things—including her most precious possessions collected throughout her life—are simply tossed in the trash. Their being thrown out like this is not a bad thing. It does not make things worse in any way. There is no reason not to do it—no reason, say, to preserve them instead. Yet it is deeply, deeply sad.

This case of the old person's things is one, also, in which something's being sad can make it seem to us *as if* there is something bad here—as anyone knows who has had to clean out the house or room of a recently deceased friend or loved one. We *know* there isn't really anything bad about these things being thrown out, but—vividly aware of the very real sadness of the situation, or momentarily in its grip—it can certainly *seem* to us as if there is. It is similar, perhaps, to the way in which a stick when placed in water can look bent, even when we know it isn't.

It is the same, of course, with disposing of one's own childhood things—old toys, clothes, papers or drawings, and so on—when it is no longer practical to hold on to all of them. To do so can be sad indeed, and our awareness of how sad this is can make it *seem* to us as if there is something *bad* about these things being thrown out (over and above any harms to us or others of our no longer having access to them). This persistent—though clearly, on reflection, mistaken—impression can sometimes lead us to hold on to such things longer than we really should.

In a similar way, I am suggesting, longtermists' awareness of the deep sadness of the Parfit/Temkin extinctions is leading them—unconsciously and wrongly—to think that there is something bad here (over and above any harms involved for the final generations). This is the root problem in their first argument for the Heart.

Why exactly are the Parfit/Temkin extinctions so sad? I think it has to do with the fact that in these cases the 'human story' is being cut short. There is still so much more for humanity yet to achieve—in the Arts and Sciences, in morality, and in our civic and political institutions—and all of this potential is being thwarted or left unrealised. There is something extraordinarily sad about the thought of human beings, who have been collectively striving for so long to improve this world—to make it more fair and just, learn more about it, and leave a better version of it for their children and later generations—and who have made

remarkable strides in these directions over many years, suddenly being snuffed out without getting to progress further.²⁴

If I'm right that the Parfit/Temkin extinctions are sad because they cut short the human story, and this explains why longtermists believe—wrongly—that these extinctions are bad (over and above the harms to the final people), then this would also explain why we do not react similarly to the other cases I gave (Utopian Extinction and Happiness Wand). In these other cases, the failure to create trillions more happy beings is *not* preventing humanity from making further progress of any kind. In Utopian Extinction, everything has been achieved already, and in Happiness Wand, the failure to wave the wand does not interfere with anything happening on Earth. This is why we do not find these events sad. This, in turn, is why there is no similar tendency to think of them as bad.

I conclude that the Parfit/Temkin cases should not lead us to accept the Heart. The right thing to say about these extinctions is not that they are bad (over and above any harms to the final generations), but merely that they are deeply, deeply sad.

3 The Argument From Miserable Beings

The longtermists' second argument for the Heart also proceeds from a widely-held premise. This is the idea that it is intrinsically *bad* to create a new *miserable* being (i.e., someone whose life will, in the end, have been so bad as to have been worth *not* living). MacAskill writes: "Imagine a life that...consists only of agony and anguish...It seems entirely obvious to me that having

²⁴ For a similar view, see Jonathan Bennett, "On Maximising Happiness", in *Obligations to Future Generations*, eds. R. I. Sikora and B. Barry (Temple University Press, 1978), 69.). For a lucid discussion of what humanity's being 'cut short' might consist in or amount to, see Jonathan Knutzen's "Unfinished Business", *Philosophers' Imprint* 23 (2023): 1 (4). See also Ord, *ibid.*, who speaks eloquently of the tragedy of the human story ending prematurely. For Ord, 'tragedy' here isn't mere sadness, but intrinsic badness.

this child would be a bad thing to do.”²⁵ Ord notes—rightly—that “almost everyone has a strong intuition that adding...lives of negative wellbeing is bad”²⁶. Longtermists conclude, for reasons of symmetry, that it must also be intrinsically *good* to create a new *happy* being. MacAskill expresses the argument like this:

If we think it’s bad to bring into existence a life of suffering, why should we not think that it’s good to bring into existence a flourishing life? I think any argument for the first claim would also be a good argument for the second.²⁷

What should we make of this argument? Some philosophers reject the view that there must be symmetry across these cases. According to them, it *can* be intrinsically bad to create a new miserable being, without it being intrinsically good to create a new happy being.²⁸

But I share the longtermist’s view that symmetry across these cases makes sense. I won’t delve here into the arguments for and against symmetry. Instead, I want to try to persuade you that—despite initial appearances—it is not intrinsically bad to create a new miserable being.

The reason why, I think, it is not intrinsically bad to create a new miserable being is that doing so does not *harm* this being. It does not harm them because it does not make them any worse off than they would otherwise have been. If they had not been created, they would never have had any welfare at all (nothing would ever have been either good or bad for them). Since creating a new miserable being does not harm this being, even one iota, it is not intrinsically bad. It does not *itself* make things any worse.

²⁵ MacAskill, *ibid.*, 180.

²⁶ Ord, *ibid.*, Appendix B.

²⁷ MacAskill, *ibid.*, 180.

²⁸ See, for example, Melinda A. Roberts, “An Asymmetry in the Ethics of Procreation”, *Philosophy Compass*. 6 (11) (2011): 765–76, and Johann Frick, “Making People Happy, Not Making Happy People: A Defense of the Asymmetry Intuition in Population Ethics” (Ph.D.) Harvard University (2014).

Of course, once such a being is *in* existence, it is a terrible thing indeed that they are suffering so. More precisely, it is terrible that they are having to live with such suffering instead of getting to live without it. *This* state of affairs—*i.e.*, *their experiencing such suffering rather than living without it*—harms them greatly, and for this reason, *it* is a very bad state of affairs indeed. But their *creation per se*, since it does not harm them at all, was not itself a bad thing.

How can their suffering be a bad thing, but not their creation, if their creation is what *caused* the suffering? It is because while their creation caused the *existence* of the suffering, it is not the bare existence of this suffering that is what's bad here. What's bad here is *that this person is suffering rather than living without such suffering*. This latter state of affairs was not caused by the creation. Creating this person did not cause them to be *in this suffering state rather than a state without such suffering*. It did not cause them, in other words, to not have a better life. All it caused was that they exist in this state *rather than never existing at all*. There was never any option (we are assuming) of creating *this very same being* in a better state.

Things are different, of course, if you create a new being and then *treat them badly*—either by causing them to suffer or failing to help them when you easily could do so. If you treat badly a new being, then your actions *do* make someone worse off than they would otherwise have been, and so *do* count as harming them. But here, the harmful actions are just your post-creation treatments, not the creation itself. It is such treatments, not the creation, that are bad.

You might object: 'But what about a case where a person creates a new being who will *certainly* be miserable, a being who *cannot* be helped in any way? Here, there are no post-creation treatments, but this person still seems to have acted terribly. So, here, it must be the act of creation itself that is bad. To flesh out the example, suppose there is a *misery wand*, which whenever it is waved, creates a new miserable being on a distant planet that nobody else will ever find out about. This wand has no other powers, including powers to improve things for this being. Waving this wand seems clearly a bad thing to do.'

We certainly do recoil at the thought of somebody waving such a wand. But, here, I think, what we are reacting to is the odious *character* of such a person. We recognise, instinctively, that only a dreadful *person* would wave such a wand. When we think ‘how terrible!’ or ‘how awful!’, this is a reaction, fundamentally, to the person, rather than to the act itself.

What could be wrong with such a person, if they are not *doing* anything bad or making the world worse in any way? One possibility is that they have an odious *intention*. Someone who would wave such a wand might *believe* (falsely) that doing so harms the newly created being. In this case, they would be a kind of sadist. Alternatively, even if they are not seeking to harm the new being, they might nonetheless be seeking to take pleasure in the existence of further suffering. This, too, would be a form of sadism and utterly odious.

More generally, somebody who would wave such a wand seems lacking or deficient in ordinary feeling or human sentiment. A good or decent person would simply *mind* the thought of waving such a misery wand, even if they believed that since waving it would not harm the newly created being, there is nothing intrinsically bad about doing so. A decent person would feel this way *simply as an offshoot or byproduct of the qualities or capacities which make them count as good or decent*. Just as (as I suggested above) somebody who understands the nature and history of humanity would feel sad at the thought of humanity’s ending prematurely even though there is no intrinsic disvalue in its so ending, so somebody who has some close friendships and who understands others well would simply balk at the thought of creating a new miserable being even though there is no disvalue in doing so. Somebody who would wave a misery wand is either lacking in understanding or affectively cold in a disturbing way.

Suppose Billy is offered by a sadist a large cash reward for waving the misery wand. He refuses, saying “I couldn’t possibly harm someone in this way.” It is then pointed out to him that waving the wand does not in fact harm the newly created being since it does not make this being any worse off than they would otherwise be. Upon realising this, Billy immediately perks up and declares: “Why, of course! How wonderful. Now I can wave this

wand and gain the cash reward without harming anyone. Everyone affected will benefit and no one will be harmed!”

Billy troubles us greatly. The reason for this, I believe, is not only that he might be enabling sadism in someone else, but that he himself seems emotionally lacking in a concerning sort of way. A person, we feel, should simply *mind* the thought of waving such a wand. It should upset or disturb them. It should not ‘sit well’ with them. This seems so *even if it is not intrinsically bad to wave the wand*. Billy, we can tell, from his eagerness to wave the wand and apparent lack of qualms or perturbation about doing so, has an impoverished emotional life and quite likely an impoverished understanding of humanity or other people.

Here’s what I’m suggesting, in a nutshell: there is no need to posit intrinsic disvalue in the act of waving the misery wand in order to explain what is wrong with *Billy*, and *Billy himself* (or his character) is sufficient to explain our revulsion at his waving it.

Thomas Hill Jr. makes a similar suggestion in a different context.²⁹ Hill asks us to consider *The Asphalter*:

A wealthy eccentric bought a house...surrounded by a beautiful display of grass, plants, and flowers, and...shaded by a huge old avocado tree. But the grass required cutting, the flowers needed tending, and the man wanted more sun. So he cut the whole lot down and covered the yard with asphalt. After all it was his property and he was not fond of plants.³⁰

According to Hill, we cannot say that this man’s action is bad or that it makes the world worse, since (we may assume) it harms no sentient beings. Instead, the right thing to say here is just that this man is a dreadful sort of person. Hill writes:

Rather than argue directly with destroyers of the environment who say, “Show me why what I am doing is immoral,” I want to ask, “What sort of person would *want* to do what they

²⁹ Thomas E. Hill Jr., “Ideals of Human Excellence and Preserving Natural Environments,” *Environmental Ethics*, 3 (1983).

³⁰ Hill, *ibid.*, 213.

propose?”...Even if there is no convincing way to show that the destructive acts are wrong (independently of human and animal use and enjoyment), we may find that the willingness to indulge in them *reflects the absence of human traits that we admire and regard morally important*.³¹

Hill’s suggestion seems exactly right to me, and a similar thing applies, I am suggesting, in the case of someone who would want or be willing to create a miserable being. The trouble is fundamentally with this person’s character, not with the act itself.

You might object: ‘But surely there is a *reason* not to create a new miserable being! It is not merely that a person who would wave the misery wand is troubling in some way. Instead, there is some consideration such that if this person were to properly understand it, this understanding might itself move them to refrain.’

I agree there is a reason not to wave the misery wand. But the reason here, fundamentally, I think, is a reason *not to be a certain sort of person*. If you are seriously considering waving the wand, or even tempted to do so, you should immediately drop everything and radically revise your entire approach to life. There are clearly things you do not understand well, or else your emotional life is seriously impoverished. In particular, you should urgently try to enter into better or closer relationships with others, in the hope that you might come to better understand them or develop a deeper or richer array of human emotions. A failure to do so is terrible for you, and bad also for those you encounter.

We can sum up as follows. If somebody creates a new miserable being, it will be a terrible thing indeed that this being is suffering rather than being better off. What’s more, the creator is clearly a dreadful person, lacking in ordinary feeling or human sentiment. Last but not least, this person had a very strong reason not to create such a being, for their priority should have been to urgently and radically change their self and whole approach to life. There is no reason to add to all this that the act of creation

³¹ Hill, *ibid.*, 217, my emphasis.

itself was intrinsically bad or made things worse in some way. Moreover, there is a reason *not* to add this: this act does no harm.

For these reasons, I believe, the longtermist's second argument for the Heart also fails. We have no good reason to conclude from it that there is intrinsic value in the creation of new happy beings.

4 Future Sentimentalism

If the Heart is false, and longtermism along with it, have we any future-regarding reasons at all—say, to prevent an asteroid striking Earth in 200 years from now, or more generally to safeguard humanity's future? Or should we instead be focused solely on improving the welfare of current people?

The first thing to say is that we clearly do have a reason to prevent an asteroid striking Earth in 200 years from now, a reason provided simply by the welfare of those people who will be alive *then*. Longtermists are right that temporal distance is morally irrelevant. If we can help people who will be alive in 200 years from now (or in a trillion years from now, for that matter), then we have a reason to do so, even at a cost to ourselves.

My agreement with longtermists goes further still. I agree with them that we have a reason not only to help future people, but to make it more likely *that there will be at least some future people in the first place*. We should indeed, just as they maintain, be trying to make it the case that humanity will survive long into the future in conditions conducive to flourishing. However, while longtermists believe that our reason here is fundamentally to add more happy beings to the world, I think it has a different basis.

To explain this basis, I want to return to the case discussed earlier of creating a new miserable being. I claimed that our reason not to create such a being is provided, not by what we would be doing *per se* (either by the nature of the act itself or by its consequences), but by what sort of a person would do such a thing. Our reason to not create such a being is ultimately a reason *to not be a certain sort of person*. Somebody who would want or be willing to create a new miserable being is clearly deficient or lacking in normal human sentiments or in their understanding of

others. If you are like *that*, you have a strong reason to immediately and radically change your whole approach to life. *This* is the basis of our reason not to create a new miserable being.

Something similar applies, I now want to suggest, when it comes to safeguarding humanity's future. If you are unmoved by thoughts of humanity's future—in particular, our continuing to make further progress in the Arts, the Sciences, morality, and our institutions—if such thoughts leave you emotionally cold, then (like somebody who wants or is willing to create a new miserable being) you clearly have an impoverished emotional life, where this is likely attributable to a narrow or limited understanding of humanity and its history. If you are like *that*, then you have an overriding reason to immediately and radically change the sort of person you are.

Somebody with a good understanding of humanity and its history—the history of our strivings to make this world more fair and just, learn more about it, and leave a better version of it for our children and later generations—would find the thought of humanity's story being cut short deeply sad or upsetting. They would simply *want* humanity to keep on going and make further progress—not because they see some intrinsic value in this, but rather as an offshoot or byproduct of the understanding they have of humanity and their emotional or sentimental nature. We have a powerful reason to be the sorts of people who are curious about things, understand things well, and feel appropriately and deeply. This reason is partly self-interested, and partly other-regarding (i.e., for the sake of others whom we will encounter in life). *This* is the basis of our reason to act to safeguard humanity's future.

Our reason to act to safeguard humanity's future, in other words, is provided not by there being great value in this end itself, but by the great value in our being the sort of people who are interested in the humanity and the world at large, understand it richly, and are moved by our thoughts of humanity's future trajectory. This view I call *Future Sentimentalism*.³²

³² It is worth distinguishing my view from Samuel Scheffler's in *Death and the Afterlife* (Oxford, 2013). Scheffler argues that current people have a strong self-interested reason to safeguard humanity's future, one

What does this mean, in practice? It means we should be pursuing policies that help educate people more about humanity and its rich history, as well as encourage them to be emotional beings, who feel fittingly and deeply in response to an appreciation of the facts. A big part of this is fostering intellectual curiosity. When people are interested in the world around them (including especially other people), they are much more likely to come to understand things better and feel accordingly.

How much, on my view, should current people be willing to sacrifice in order to safeguard the future of humanity? This will be determined by how much those with a rich understanding of humanity and its history and who are feeling appropriately and deeply would *want* or be *willing* to give up. I suspect they would be willing to give up quite a lot. Note that their goal here would be, not to further populate the distant future, but to ensure that humanity carries on long into the future, making further progress in the Arts, the Sciences, morality, and our institutions.

5 An Objection

It might be objected that even if I have shown that we do not have sufficient evidence to believe the Heart, we should still be

grounded in the fact that so many of the activities that give meaning to our lives depend on our expectation that humanity will continue. Without such an expectation, many of these activities would no longer make sense, and we might come to feel unhappy or depressed. On such a view, our reason to safeguard humanity's future is that it is very good for us to believe that humanity has a future. On my view, by contrast, our reason is not some beneficial consequence of our believing that humanity has a bright future. Rather, it is a reason to be people who are curious or interested in the world around them, understand things well, and feel appropriately and deeply. People like that will, as a matter of fact—as an offshoot or byproduct of these traits—care about the future and want to take action to safeguard it. Note that Scheffler himself believes in the existence also of what he calls “attachment-independent” reasons to safeguard humanity's future—reasons that are provided by the good we would be doing by so acting. This point is sometimes overlooked by longtermists. For more details, see his *Why Worry About Future Generations?* (Oxford, 2018)

longtermists, since the Heart still *might* be true. All longtermists need, it might be said, is that there is some non-negligible chance that the Heart is true. Given the *huge* amount of value there would be in further populating the distance future *if* the Heart is true, this chance is enough for it to be *subjectively* rational (i.e., rational, given our limited epistemic state) for us to make big sacrifices now in order to further populate the distant future.

My response to this objection is that while I am by no means *sure* that the Heart is false, it seems to me almost certainly false. Not only do the two arguments for it I have considered in this paper fail to establish its truth, they seem to provide little or no evidence for it. There are, as I have argued, much better explanations available of our feelings about human extinction and the creation of new miserable beings than those given by longtermists. What's more, there is a very powerful reason to reject the Heart: creating a new happy being does not benefit this being.

Suppose someone says: "But shouldn't we wave the happiness wand (from Section 2) *on the off chance* that there is intrinsic value in adding new happy beings, given the truly massive number of new happy beings it would create?" No. The probability that the Heart is true is simply too low. It is so low as to not be worth taking into account in our decision-making at all.

What of the fact that so many highly intelligent people find the Heart plausible? Shouldn't we attach some weight to this fact?

It remains to be seen that these people will still find the Heart plausible after considering the arguments of this paper. In any case, intelligent people have been wrong before, even when many are in agreement. In the case at hand, there is a good explanation available of where these intelligent people are going wrong.

In particular, I hope that longtermists will see, after reading this paper, that many of the things they want to insist on—that there is a sense in which the Parfit/Temkin extinctions are deeply tragic, that there is something awful about a person's deciding to create a new miserable being, that we have strong reasons indeed not only to help future people, but to ensure that there will be at least some of them in the first place—are things we can

consistently maintain and readily explain even if the Heart is false.

6 Conclusion

In this paper, I have tried to explain what is wrong with longtermism. The trouble with it, I've argued, is with the idea located at its very heart, that there is intrinsic value in adding new happy beings. There is no such value, since creating new happy beings does not benefit these newly created beings.

Sacrificing the interests of current people in order to further populate the distant future is a bad idea because it is to forgo helping some people in order to help nobody.

Let's focus our attentions on current people. In particular, let's help them to better understand the world around them, including especially each other, the value of what we have created here on Earth, and the history of our collective strivings. If we do this, there is a very good chance that most of us will become the sorts of people who will desperately *want* to see humanity survive long into the future, continuing to make progress in the Arts, the Sciences, morality, and our institutions.