

# The toxic ideology of longtermism

Alice Crary

The intellectual movement that calls itself longtermism is an outgrowth of Effective Altruism (EA), a utilitarianism-inspired philanthropic programme founded just over a decade ago by Oxford philosophers Toby Ord and William MacAskill. EA, which claims to guide charitable giving to do the ‘most good’ per expenditure of time or money, originally focused on mitigating the effects of poverty in the global South and the treatment of animals in factory farms.<sup>1</sup> This initially modestly-funded, Oxford-based enterprise soon had satellites in the UK, US, and elsewhere in the world, several of which became multi-million-dollar organisations, while the amount of money directed by EA-affiliated groups swelled to over four hundred million dollars annually, with pledges in the tens of billions.<sup>2</sup> During this period, Ord and MacAskill starting using the term ‘longtermism’ to mark a view championed by members of a conspicuous subset of effective altruists, many affiliated with Oxford University’s Future of Humanity Institute. The view is that humanity is at a crossroads at which we may either self-destruct or realise a glorious future, and that we should prioritise responding to threats to the continued existence of human civilisation. The ‘existential risks’ – to use the term introduced by Oxford philosopher and Future of Humanity Institute founder Nick Bostrom<sup>3</sup> – that longtermists rank as most probable are AI unaligned with liberal values and deadly engineered pathogens. They urge us to combat these risks to make it likelier that humans (or our digitally intelligent descendants) will live on for millions, billions, or even trillions of years, surviving until long after the sun has vaporised the earth, by colonising exoplanets.

Ord published a monograph defending a longtermist stance in early 2020, and MacAskill followed suit in the summer of 2022.<sup>4</sup> Ord’s book received plaudits in high-profile venues,<sup>5</sup> and MacAskill’s was a best-seller

that came with a blitz of largely positive media attention, including a *New Yorker* profile, a review featured on the cover of *Time*, and an appearance on *The Daily Show*, as well as an endorsement from Elon Musk.<sup>6</sup> This was the coming out party for a tradition that, despite its notable influence in Silicon Valley and elite universities, had previously flown mostly under the radar.

The public mood changed in mid-November of 2022, when one of the movement’s biggest funders, the crypto exchange FTX, declared bankruptcy. It was then known that MacAskill and FTX’s CEO Sam Bankman-Fried had been acquainted since 2012, when MacAskill advised Bankman-Fried, at the time an MIT undergraduate, to channel his altruistic zeal into ‘earning to give’.<sup>7</sup> It was also known that, with a group of Oxford-affiliated longtermists, MacAskill had been an advisor to FTX’s charitable Future Fund, and that the Future Fund had committed large sums to building EA’s own institutions, including fourteen million dollars to MacAskill’s main organisation, the Centre for Effective Altruism, fifteen million to Longview Philanthropy, for which MacAskill is an advisor, and another roughly seven million to fellowships, prizes, and the like at these and other organisations with which MacAskill is affiliated.<sup>8</sup> Such institutional ties have been mentioned, alongside facts about how prominent tech multi-millionaires and billionaires support longtermist projects,<sup>9</sup> in a journalistic narrative that faults longtermism in moral terms for enriching itself by indulging the self-aggrandising, techno-utopian fantasies of its donors while ignoring questions about the sources of their wealth.

This critique of longtermism is correct as far as it goes. It is also desperately incomplete. One thing it fails to capture is that an uncritical attitude toward existing political and economic institutions is part of longtermism’s philosophical DNA. The point of departure for

longtermism is EA, and, like other utilitarianism-inspired doctrines, EA veers towards forms of welfarism that are unthreatening to the status quo. This posture increasingly exposed EA to corruption during its growth into a broad-scale philanthropic movement. EA shares the tendency of large charitable foundations to undemocratically organise entire realms of public engagement, diverting money and other resources from movements for liberating social change. And it owes its ability to secure the funding requisite for this role to its affinity with political and economic systems generative of the suffering it claims to address.<sup>10</sup>

Longtermism's sins are different and more ominous, but there are points of convergence. Longtermism deflects from EA's wonted attention to current human and animal suffering. It defends in its place a concern for the wellbeing of the potentially trillions of humans who will live in the long-term future, and, taking the sheer number of prospective people to drown out current moral problems, exhorts us to regard threats to humanity's continuation as a moral priority, if not *the* moral priority.<sup>11</sup> This makes longtermists shockingly dismissive of 'non-existential' hazards that may result in the suffering and death of huge numbers in the short term if, as they see it, there is a reasonable probability that the hazards are consistent with the possibility of a far greater number of humans going on to flourish in the long term.

When longtermists turn to existential hazards, they discuss wholly natural threats (such as large asteroids hurtling toward the earth, super-volcanic eruptions, and stellar explosions) while focusing on human-caused risks, which they regard as more likely to rise to extinction-level. Alongside value-divergent AI and human-produced pathogens, they consider climate change, other forms of environmental degradation, and all-out nuclear war, and they set out to calculate the probability that these different anthropogenic threats will instigate existential disasters. This accent on existential dangers is theoretically unjustified and morally damaging, but even stripped of it, longtermism is a poor guide to solicitude for prospective humans.

Longtermism calls on us to safeguard humanity's future in a manner that both diverts attention from current misery and leaves harmful socioeconomic structures critically unexamined. As a movement, it has enjoyed stunning financial success and clout. But its success is

not due to the quality of its conception of morality, which builds questionably on EA's. Rather, it is due to longtermism's compatibility with the very socioeconomic arrangements that have led us to the brink of the kinds of catastrophes it claims to be staving off. At issue is not only an especially dangerous, future-facing variation on ideologies, like EA, that thwart struggles for liberating change with suggestions of the cure-all properties of existing economic tools. It is a variation lacking any plausible rationale, since many of these struggles have long contributed to the area longtermism wrongly represents as its innovation – fighting for a just and livable future.

The longtermist enterprise has been publicly thrashed for its ties to FTX, but it remains well-funded and well-positioned to repair its reputation and go on enlisting earnest individuals to energetically support and spread it. There is a pressing need to criticise its theoretical weaknesses and forcefully bring out its material harms, exposing it as the toxic ideology it is.

## Longtermist moral logic

The ethical core of longtermism is a set of commitments, shared with EA, from the moral tradition of consequentialism. For consequentialists, the mark of right action is producing outcomes that are best in the sense of containing the greatest amount of value. That leaves open what is of value, and, although longtermists often insist on respect for uncertainty about the correctness of any one moral theory, they still incline toward versions of consequentialism that identify value with wellbeing and so fall under the heading of utilitarianism. In making these theoretical moves, longtermists help themselves to a methodological assumption that is itself morally significant. Together with effective altruists and many others partial to utilitarian stances, they assume that wellbeing is discernible from a dispassionate and abstract 'point of view of the universe'.<sup>12</sup> That is morally significant because it is what seems to make it possible to use wellbeing as a measure for comparing outcomes anywhere – not only across space to the global poor and across species to non-human animals but also across time to those living in the far distant future.

Longtermism proper emerges from within a set of contemporary ethical discussions, typically described

as composing the field of ‘population ethics’, in which utilitarianism-tinged modes of thought are applied to prospective humans.<sup>13</sup> Debates among population ethicists pivot around questions about whether our moral assessments appeal to total aggregate wellbeing, average wellbeing, or wellbeing above a certain critical level, as well as around questions about whether moral assessments reflect equal versus unequal distributions of wellbeing. A signature gesture of these moral theorists is insisting that their research programme is extremely difficult, presenting participants with nearly intractably vexing problems.<sup>14</sup> But the issues that trouble population ethicists presuppose their methodologically abstract, calculative approach to people and circumstances. Their conundrums don’t arise for moral thinkers who reject this method as unsuited to the subject matter.

What distinguishes longtermism from other positions within population ethics is a pair of related claims, one empirical and the other ethical. The empirical claim is that we live ‘at a time uniquely important to humanity’s future’ in which ‘major transitions in human history have enhanced our power and enabled us to make extraordinary progress’ while also putting us at risk of self-annihilation.<sup>15</sup> The ethical claim has to do with what population ethicists call ‘the intuition of neutrality’, that is, the intuition that what matters morally is the quality of peoples’ lives, not how many people there are. Thinkers who incline toward neutrality hold that whether a greater or smaller number of people live at a given time is in itself morally neutral. Longtermists in contrast reject this notion of neutrality, maintaining that any additional person who lives makes the world better, as long as the person enjoys adequate wellbeing.

This is the ethical backdrop against which longtermists’ empirical claim about humanity standing at a historical ‘precipice’, a time both of great promise and of increased risk of auto-extinction, seems momentous. Now it appears that a circumstance in which human beings die out in a few thousand years is worse, by many orders of magnitude, than one in which trillions of humans live on to flourish in the distant future. It appears that it would be a massive moral achievement to improve the prospect of avoiding extinction by even a fraction of a percentage. The endeavour would be so important that it would justify almost any means, however seemingly callous or appalling, including steps that resulted in the near-term

suffering and death of millions.<sup>16</sup> Not that all longtermists explicitly contemplate extreme or violent actions to avoid existential disasters.<sup>17</sup> Even those who actively oppose such measures, however, offer frighteningly few safeguards to keep their moral calculations from echoing the reasoning of murderous dictators and sci-fi villains.

## Empty ethical equations

Longtermists’ turn to existential risk marks a dramatic shift from the concern with present and near-term suffering that is the hallmark of their effective altruist progenitors. Unsurprisingly, some advocates of EA are fiercely critical of longtermism. That includes Peter Singer, whose contributions to utilitarian ethics were EA’s original inspiration. Singer is skeptical about whether humanity is indeed at a uniquely portentous moment in history, and he de-emphasises existential risk in a manner that indicates impatience with longtermists’ commitment to the posture they call non-neutrality. His aim is to redirect attention back to EA’s accent on suffering now and in the short-term. ‘If we are at the hinge of history’, he writes, ‘enabling people to escape poverty and get an education is as likely to move things in the right direction as almost anything else we might do; and if we are not at that critical point, it will have been a good thing to do anyway’.<sup>18</sup>

Singer proposes to strip longtermism of the claims that differentiate it from EA, leaving a future-oriented outlook that might be described as a generic position within population ethics. Such a future-directed EA would, he suggests, be an authoritative guide to doing good for human beings to come. But this suggestion reflects a fundamentally limited diagnosis of what ails longtermism. Even without the claims that lead its advocates to wrongly represent existential risks as swamping other moral concerns, the tradition is incapable of furnishing an understanding of our social circumstances that could responsibly inform future-oriented action.

The grounds for this more negative appraisal of longtermism can be found in one of the most well-known critiques of EA. Since EA’s inception, critics have noted that its emphasis has been on assessing single action-types (e.g., medical, public health, or educational interventions) in terms of the sort of wellbeing grasped by the metrics of welfare economics. They have observed that

EA's slant toward welfarism is at the same time a slant away from questions of justice, and they have revisited in reference to EA a classic charge against utilitarianism. The charge's thrust is that EA is politically corrosive because it neglects the structural roots of global misery and so weakens political bodies capable of challenging those structures, ensuring the regular reproduction of suffering.<sup>19</sup>

Some effective altruists respond to this critique by arguing that, even if EA has in practice veered toward welfarism, there is in principle nothing to keep it from evaluating social movements' coordinated efforts to fight for more just social arrangements and also nothing to prevent it from using the kinds of qualitative metrics that we find in disciplines in the social sciences, such as sociology and political theory.<sup>20</sup> But this rejoinder falls flat. It is undercut by effective altruists' reliance on the god's eye moral method that seems to enable them to quantify values across space and species and arrive at aggregative judgments of 'most good'.

This methodological stance disqualifies anyone who adopts it from discerning the systematic injustices targeted by social justice movements. When participants in anti-racist, feminist, and Indigenous rights movements protest sustained physical, psychological, and political violence against specific oppressed human groups, they are moved by structural obstacles to flourishing and the absence of reparations. It is not possible to adequately grasp the nature of such wrongs without an appreciation of the history and function of the social mechanisms that reproduce them. Attempts to understand these injustices, when approached in the abstract and aperspectival manner characteristic of EA, uninformed by pertinent historical, cultural, and political considerations, are bound to misfire. They also risk strengthening the oppressive structures in question because one way in which these structures function is by obscuring the historically and socially specific suffering of the oppressed. That, then, is why EA is unable to slough off the allegation that it has a politically conservative, welfarist bent. It lacks the immanent resources necessary for illuminating systematic injustice and envisioning appropriate remedies to it.<sup>21</sup>

Longtermism is tainted by the same lack. Population ethics, the original home of longtermism, is premised on the assumption that, appropriately specified, an abstract account of an action's effects on the lifetime wellbeing of

prospective populations equips us to answer questions about the action's rightness. This assumption is false. Instead of making it possible to determine what counts as right action, a detached approach obscures from view just and unjust relationships that are part of these determinations' lifeblood. Because the calculative enterprise in which population ethicists are engaged is based on false presuppositions, the technical headaches with which it presents them are at bottom self-inflicted injuries. The correct attitude to their disciplinary puzzles is to dissolve not solve them, and this applies to the debate, central for those population ethicists who self-denominate as longtermist, about whether to affirm an 'intuition of neutrality'. The debate's conceit is that there is a coherent abstract question about whether creating more happy people is a moral gain. But longtermists' assertion of non-neutrality is nothing more than an empty gesture. The emptiness extends to the non-neutrality-based computations that seem to support longtermism's insistence on regarding existential risk as a great or even overwhelming moral priority. Longtermists' distinctive moral math simply falls apart.<sup>22</sup>

## Exploiting existential angst

This isn't yet an adequate inventory of longtermism's major weaknesses. Once we set aside the morally free-floating calculations on which longtermists build their case for an extreme prioritising of existential risks, it might seem that we retain the makings of a helpfully future-oriented practical programme. That is the gist of Singer's proposal for exchanging longtermism's fixation on historical precipices with a forward-looking, utilitarian-themed project that recentres longtermism's origins in EA. But this is a non-starter. It fails to register that EA itself is incapable of shedding light on unjust and harmful social structures or assessing efforts to resist them. Even shorn of its wrongheaded stress on existential hazards, longtermism is a treacherous guide to acting responsibly towards those who will come after us.

This emerges concretely in MacAskill's and Ord's treatments, in their respective recent books, of climate change and other forms of anthropogenic environmental destruction. Both discussions are distorted by a disturbing interest in existential risk that makes it seem per-



continent to investigate whether global heating will lead to human extinction or whether it will 'only' kill billions of humans and trillions of animals and devastate ecosystems, while still permitting the survival and ultimate flourishing of small human groups.<sup>23</sup> This misguided preoccupation with existential dangers is closely tied to other outrages, such as MacAskill's selective and highly contentious appeal to climate science in support of his chillingly casual 'best guess' that some human beings would survive 'fifteen degrees of warming'.<sup>24</sup> But, even apart from their morally disastrous hang-ups with human extinction, MacAskill's and Ord's reflections on the environmental crisis are ruinously wrongheaded. When they consider strategies for reducing greenhouse gas emissions to combat the devastation of climate change, they limit themselves to strategies that can be pursued within existing socioeconomic arrangements. This includes technological innovations such as 'clean' or low-carbon energy sources and different forms of geoengineering.<sup>25</sup> It also includes policies such as internationally

coordinated emission-reduction schemes.<sup>26</sup> MacAskill at one juncture mentions youth activism admiringly, but his point about it is simply that it can increase public support for climate pledges.<sup>27</sup> Nowhere in Ord's or MacAskill's remarks is there any real acknowledgment of the reality, repugnant to members of the billionaire class they assiduously and successfully cultivate, that meaningful environmental action will need to involve new values and substantial social change.<sup>28</sup>

Still more striking, perhaps, is that MacAskill and Ord try to diminish our sense of the urgency of environmental issues, arguing that we should regard renegade AI and human-developed pathogens as more critical because likelier to trigger human extinction. This line of argument, common among longtermists, is a further expression of the warping moral effects of a fascination with extinction risks, which seems to speak for downgrading the exigency of things that don't extinguish human life altogether, and so supports treating as relatively morally insignificant the terrible fact that huge numbers

of people are already dying, being uprooted from their communities, and suffering other great hardships because of climate change.<sup>29</sup> Yet, even within the context of MacAskill's and Ord's extinction-focused programme, it is not clear why the environment fails to loom larger. Ord argues that environmental degradation is relatively unlikely to directly produce an extinction event and more likely to generate forms of political instability that indirectly lead to one, providing the conditions for other anthropogenic dangers.<sup>30</sup> It's not clear why that should make it less imperative to attend to environmental factors, or why a deviant robot takeover should be a bigger priority, unless it's just that, considered in isolation, deviant AI appears to be a hazard addressable with the kinds of instruments that Ord and other longtermists have at their disposal. Here the drive to downplay the seriousness of environmental crisis plainly outruns the grounds for doing so.

Longtermism is marred not only, therefore, by a misjudged positioning in population ethics that swings it toward existential risk but also by methodological presuppositions that prevent it from recognising that movements for social change, such as the environmental movement in its interplay with anti-racist and other social justice movements, have long been engaged in the kind of future-facing social enterprise it preposterously credits itself with inaugurating.<sup>31</sup>

These objections are not at base about the troubling fact that the tradition is the brainchild of a group of white men at an elite university, some of whom have records of racist statements.<sup>32</sup> More fateful is a dimension of longtermism's signature theories of existential risk. These theories treat as less urgent those anthropogenic hazards that won't snuff out humanity altogether, and the theories' adherents place the currently intensifying human-caused climate crisis squarely in this category, encouraging us to regard as morally less important the suffering and death it is occasioning. The harms in question are falling in dramatically lopsided fashion on racialised and Indigenous groups the world over, groups whose very vulnerability to these harms is a product of long histories of injustice. Such theory-induced callousness to losses and damages visited grossly unequally on racialised people licenses talk of a racist strain in longtermist thinking, and individual longtermists deepen this strain in specific ways. A well-placed young longtermist once ar-

gued that inhabitants of rich countries are generally more 'innovative' and 'economically productive' and that saving their lives is hence substantially more important for humanity's future than saving lives in poor countries.<sup>33</sup> Today some of the tradition's most prominent champions advocate projects of bio-enhancement, reminiscent of twentieth-century eugenics, aimed at developing a transhuman species that is better equipped for survival in the long-term.<sup>34</sup> These sorts of reinforcements of longtermism's racist streak are only strengthened by the tradition's inability to grasp, and consequent proclivity to make invisible, contributions to revolutionary anti-racist struggle.<sup>35</sup>

## Mega-philanthropic delusions

The story of longtermism is not just a tale of a no good, very bad moral theory. As the coffers of longtermism's institutes and related charities have swelled, it has begun to enact its priorities, funding research on misaligned AI and anthropogenic pathogens and supporting institution-building, with research grants as well as grants to EA's and longtermism's institutes.<sup>36</sup> Its arrival as a philanthropic player exposes it to concerns about having an unmerited sway on social issues. Like other wealthy private foundations, longtermist organisations are able to specify what counts as good and shape civic life without real public answerability. In the US and elsewhere, tax exemptions of well-funded private charities take from the public till huge sums that voters could otherwise have directly determined how to spend, and, apart from relatively insignificant tax obligations and reporting duties, there is little accountability. This is a money-fueled arrangement involving 'the exercise of wealth-derived power in the public sphere with minimal democratic controls and civic obligations'.<sup>37</sup> With its growth into a movement, longtermism has joined this undemocratic commandeering of the public realm, using its financial heft to promote its dangerous obsession with existential risk.

Longtermism's moral case for accenting such risk deflects from present suffering in a manner that simultaneously absolves harmful socioeconomic mechanisms from criticism and hastens the sorts of hazards it is supposed to head off. Yet it has been singularly successful at attracting rich backers to its project. In treat-

ing the economic arena to which these individuals owe their wealth as critically off limits, it positions them to look upon themselves, not as complicit in the arena's injustices, but as singled out by their success in it to be world saviours.<sup>38</sup> A deceitful narrative of selfless heroes riding to humanity's rescue has proven ideologically effective, and it seems clear that many longtermists – students, researchers and members of the public, as well as donors – are sincerely committed to what they take to be a uniquely important moral enterprise. But their sincerity is no argument against the corruption of a movement that uses a bankrupt morality to justify profiting from the systems most threatening to the future it claims to secure.

The fact that some major supporters of longtermism, such as Bankman-Fried, have been suspected of financial fraud is a sideshow to the main event. Longtermism's corruption is inseparable from the way in which its core ideas are put into practice, and the baseness is still there when its programmes are pursued with rigorous legality. A critique of longtermism that enabled its adherents to see it in this harshly revealing light would be a welcome step towards envisioning and enacting a just and livable future.<sup>39</sup>

*Alice Crary is a philosopher and writer based at The New School who divides her time between New York, Oxford, and Berlin. Her most recent book, co-edited with Carol Adams and Lori Gruen, is The Good it Promises, The Harm it Does: Critical Essays on Effective Altruism (2023).*

## Notes

**1.** In a previous article for this journal, I criticised EA with particular reference to its tendency to work against its own commitment to the cause of non-human animals. See 'Against 'Effective Altruism'', *Radical Philosophy* 2.10 (Summer 2021), 33–43.

**2.** Benjamin Todd, co-founder of the EA-affiliate 80,000 hours, estimated in summer 2021 that total pledges to EA had reached forty-six billion dollars. See 'Is effective altruism growing? An update on the stock of funding versus people', July 21, 2021, at <https://80000hours.org/2021/07/effective-altruism-growing/>. In a post in the EA Forum on May 9, 2022, MacAskill gave a more conservative estimate of thirty billion. See <https://forum.effectivealtruism.org/posts/cfdnJ3sDbCSkShiSZ/ea-and-the-current-funding-situation#fnaavi420x9rk>.

**3.** See Nick Bostrom, 'Existential Risks: Analyzing Human Extinction Scenarios and Related Hazards', *Journal of Evolution and Technology* 9:1 (2002).

**4.** Toby Ord, *The Precipice: Existential Risk and the Future of Humanity* (New York: Hachette Books, 2020); and William MacAskill *What We Owe the Future* (New York: Basic Books 2022).

**5.** See, for example, Jim Holt, 'The Power of Catastrophic Thinking', *The New York Review* (February 25, 2021); and also the mention of Ord's book in the 'Briefly Noted' section of *The New Yorker* (April 5, 2020).

**6.** See Gideo Lewis-Kraus, 'The Reluctant Prophet of Effective Altruism', *The New Yorker* (August 8, 2022), <https://www.newyorker.com/magazine/2022/08/15/the-reluctant-prophet-of-effective-altruism>; and Naina Bajekal, 'Want to do More Good? This Movement Might Have the Answer', *Time Magazine* (August 10, 2022). MacAskill appeared on *The Daily Show with Trevor Noah* on September 27, 2022, <https://www.cc.com/video/8fl6g9/the-daily-show-with-trevor-noah-william-macaskill-what-we-owe-the-future>; and on August 2, 2022, Musk retweeted MacAskill's book announcement to his own more than 120 million followers, commenting: 'Worth reading. This is a close match for my philosophy', <https://twitter.com/elonmusk/status/1554335028313718784>.

**7.** For one of the most detailed accounts of MacAskill's acquaintance with Bankman-Fried, see Adam Fisher, 'Sam Bankman-Fried has a Savior Complex – and Maybe You Should Too', *Sequoia* (September 22, 2022), <https://web.archive.org/web/20221027180943/https://www.sequoiacap.com/article/sam-bankman-fried-spotlight/>.

**8.** For these figures, see John Hyatt, 'Sam Bankman-Fried's Donations to Effective Altruism Nonprofits Tied to an Oxford Professor are at Risk of Being Clawed Back', *Forbes* (November 17, 2022), <https://www.forbes.com/sites/johnhyatt/2022/11/17/disgraced-crypto-trader-sam-bankman-fried-was-a-big-backer-of-effective-altruism-now-that-movement-has-a-big-black-eye/?sh=6c346564ce78>. Although, together with four other EA-affiliated advisors to FTX's Future Fund, MacAskill stepped down from his role the day before FTX's bankruptcy filing (see <https://forum.effectivealtruism.org/posts/xafpj3on76uRDoBja/the-ftx-future-fund-team-has-resigned-1>), it seems clear that alarms about Bankman-Fried's conduct were raised by effective altruists not days but years before the FTX debacle. See Reed Albergotti and Louise Matsakis, 'Effective Altruism Group Debated Sam Bankman-Fried's Ethics in 2018', *Semafor* (November 18, 2022), <https://www.semafor.com/article/11/18/2022/effective-altruism-group-debated-sam-bankman-frieds-ethics-in-2018>.

- 9.** This includes, for example, both Facebook co-founder Dustin Moskovitz, who co-founded the grantmaking foundation Open Philanthropy that takes longtermism as one of its main cause areas and Skype founder Jaan Tallinn, who co-founded Cambridge University's longtermism-oriented Centre for the Study of Existential Risk.
- 10.** For development of these charges against EA, see Carol Adams, Alice Crary and Lori Gruen, eds., *The Good It Promises, The Harm It Does: Critical Essays on Effective Altruism* (Oxford: Oxford University Press, 2023).
- 11.** Longtermists distinguish weak versions of their creed, which treat existential risk as a moral priority, and strong versions, which treat it as the moral priority. Some, such as MacAskill, defend the weaker doctrine in public-facing work (e.g., *What We Owe the Future*) while championing the stronger one in scholarly writing (e.g., 'The Case for Strong Longtermism', co-authored with Hilary Greaves, <https://globalprioritiesinstitute.org/wp-content/uploads/The-Case-for-Strong-Longtermism-GPI-Working-Paper-June-2021-2-2.pdf>.) The argument of the current article does not distinguish weak and strong longtermism and bears on both.
- 12.** This phrase was introduced by nineteenth-century utilitarian Henry Sidgwick and adopted by contemporary utilitarian Peter Singer, whose work, discussed below, is foundational for EA. See Singer and Katarzyna de Lazari-Radek, *The Point of View of the Universe: Sidgwick and Contemporary Ethics* (Oxford, Oxford University Press, 2014). Not all effective altruists and longtermists use this nomenclature, but all make moves in value theory that treat moral thought as coming from an Archimedean point.
- 13.** The idea of population ethics (although not the label) comes from Derek Parfit's *Reasons and Persons* (Oxford, Oxford University Press, 1984).
- 14.** For MacAskill's version of this gesture, see *What We Owe the Future*, 169–170.
- 15.** The inset quote is from Ord, *The Precipice*, 11.
- 16.** One of the most vocal critics of longtermism, Émile Torres, has helpfully stressed this aspect of the tradition's moral logic. See 'Against Longtermism', *Aeon* (October 19, 2021), <https://aeon.co/essays/why-longtermism-is-the-worlds-most-dangerous-secular-credo>.
- 17.** But see Nick Bostrom, 'The Vulnerable World Hypothesis', *Global Policy* 10:4 (2019), 455–476.
- 18.** Singer, 'The Hinge of History', *Project Syndicate* (October 8, 2021), <https://www.project-syndicate.org/commentary/ethical-implications-of-focusing-on-extinction-risk-by-peter-singer-2021-10?barrier=accesspaylog>.
- 19.** Some of the earliest versions of this criticism of EA, later dubbed 'the institutional critique' of EA, were presented as responses to a 2015 forum in the *Boston Review* on Peter Singer's 'The Logic of Effective Altruism', <https://www.bostonreview.net/forum/peter-singer-logic-effective-altruism/>. See especially the responses by Angus Deaton ([https://www.bostonreview.net/forum\\_response/response-angus-deaton/](https://www.bostonreview.net/forum_response/response-angus-deaton/)) and Iason Gabriel ([https://www.bostonreview.net/forum\\_response/response-iason-gabriel/](https://www.bostonreview.net/forum_response/response-iason-gabriel/)).
- 20.** For a clear defense of EA along these lines, see Jeff Sebo and Peter Singer, 'Activism', in Lori Gruen, ed., *Critical Terms in Animal Studies* (Chicago: University of Chicago Press, 2018), 33–46.
- 21.** This paragraph rehearses in compact form the main – 'composite' – critique of EA that I develop in 'Against "Effective Altruism"'.
- 22.** I owe the ideal of 'longtermist moral math' to Kieran Setiya's critique of MacAskill in 'The New Moral Mathematics', *The Boston Review* (August 15, 2022), <http://www.bostonreview.net/articles/the-new-moral-mathematics/>. Setiya's critique, one of the best to date, is insightful, though not critical enough. It falls short in treating MacAskill's longtermist theory as a mere set of ideas as opposed to a materially significant ideology.
- 23.** See Ord, *The Precipice*, Chapter 4, and MacAskill, *What We Owe the Future*, Chapter 6.
- 24.** MacAskill, *ibid.*, 137. See also Ord's claim, in *The Precipice*, 110, that thirteen degrees of warming would be 'a global calamity of an unprecedented scale' but not an existential catastrophe.
- 25.** See Ord, *The Precipice*, 112–113, and MacAskill, *What We Owe the Future*, 135.
- 26.** MacAskill, 135.
- 27.** *Ibid.*
- 28.** Late in his book, MacAskill surprises by saying he advocates 'systemic change'. 'In order to solve climate change', he writes, 'what we actually need' is not 'personal consumption decisions' but for 'companies like Shell to go out of business' (*ibid.*, 232). For this, he recommends donations to 'effective' non-profits, presenting his 2016 book *Doing Good Better* (London: Penguin, 2016) as a guide. This recommendation undermines his avowed system-changing aims, since *Doing Good Better* is a welfarism-oriented EA manifesto with a conservative bent that lacks any serious critical engagement with anthropogenic global heating. (See Rupert Read, 'Must Do Better', *Radical Philosophy* 2.01 (February 2018).) MacAskill's talk of systemic change in *What We Owe the Future* is empty rhetoric, disconnected from his practical proposals and commitments. That hasn't stopped it from fooling some commentators. In 'An Effective Altruist? A



philosopher's guide to the long-term threats to humanity', *Times Literary Supplement* (September 9, 2022), 9–11, Regina Rini breezily, and wrongly, cites this passage as evidence that MacAskill is 'no corporate shill' (9).

**29.** There is almost no acknowledgement of these harms in MacAskill's and Ord's recent books. On page 136 of *What We Owe the Future*, MacAskill does consider the prospect of global heating doing great damage to poorer, agrarian countries in the tropics 'that have contributed the least to climate change'. But he represents this 'colossal injustice' as something that may happen in the future and simply sets aside the question of how to respond to it.

**30.** Ord summarises his ranking of existential risks in Table 6.1 of *The Precipice*. MacAskill's similar ranking is reflected in the order of treatment of risks in *What We Owe the Future*.

**31.** For MacAskill's farcical suggestion that longtermism introduces a long-neglected future-orientation to social thought, see *What We Owe the Future*, 9, where he describes 'previous social justice movements, such as those for civil rights and women's suffrage' that have 'sought to give greater recognition and influence to disempowered members of society, adding that he sees longtermism, with its concern with future people, 'as an extension of these ideals'.

**32.** On January 9 2023, Future of Humanity Institute founder and longtermist Nick Bostrom posted online about an explicitly anti-Black email he wrote in the mid 1990s, apologising for the email but doing so in a manner that is itself racist. (His easily findable post is intentionally not included here.)

**33.** Nick Beckstead, currently a research fellow at the Future of Humanity Institute, defends this view in his 2013 PhD thesis, 'On the Overwhelming Importance of Shaping the Far Future'. Beckstead has appended a note to the relevant passage of his thesis (see page 11 of <https://drive.google.com/file/d/0B8P94pg6WYCIc0IXSUVVYS1BnMkE/view?resourcekey=0-nk6wM1QIPi0qWVh2z9FG4Q>), denying that he thinks 'lives in rich countries are intrinsically more valuable' and insisting that 'it is generally best for public

health to prioritize worse-off countries'. But he hasn't disavowed the longtermist reasoning that led him to his startling early view.

**34.** Bostrom is among the high-profile longtermists who hold that a 'transformative change of human biological nature' may be key to avoiding existential catastrophe ('Existential Risk as a Global Priority', *Global Policy* 4:1 (2013), 15–31). Ord sympathises with this view. See, e.g., his claim that 'forever preserving humanity as it is now may...squander our legacy' (*The Precipice*, 239).

**35.** Longtermism suffers from serious defects beyond those discussed in this article. Its account of how non-human animals figure in future-oriented moral thought is particularly objectionable. For a compact treatment of this topic, see Carol Adams, Alice Crary and Lori Gruen, 'Coda - Effective Altruism and Future Humans' in Adams, Crary and Gruen, eds., *The Good It Promises, The Harm It Does*.

**36.** Online reports of the 2022 grants of, e.g., Longview Philanthropy, the FTX Future Fund (pre-collapse), and the longtermism-wing of Open Philanthropy reveal general alignment with the longtermist agenda of MacAskill's and Ord's books, as described in this article.

**37.** Joanne Barkan, 'Plutocrats at Work: How Big Philanthropy Undermines Democracy', *Dissent* (Fall 2013), <https://www.dissentmagazine.org/article/plutocrats-at-work-how-big-philanthropy-undermines-democracy>. Barkan's critique of mega-philanthropy belongs to a small and valuable corpus that is reprising, with reference to the Gates Foundation and today's other biggest charitable organisations, themes of a twentieth-century debate about damaging political effects of the Ford, Rockefeller, and Carnegie foundations.

**38.** For a virtuoso filmic expression of this false but alluring trope of the mega-wealthy individual as guardian of humanity, see the billionaire businessman and inventor Peter Isherwell in Adam McKay's 2021 film *Don't Look Up*.

**39.** While writing this article, I benefitted from helpful correspondence with Carol Adams, Jay Bernstein, Victoria Browne, David Cunningham, Lori Gruen, Émile P. Torres and Nathaniel Hupert.