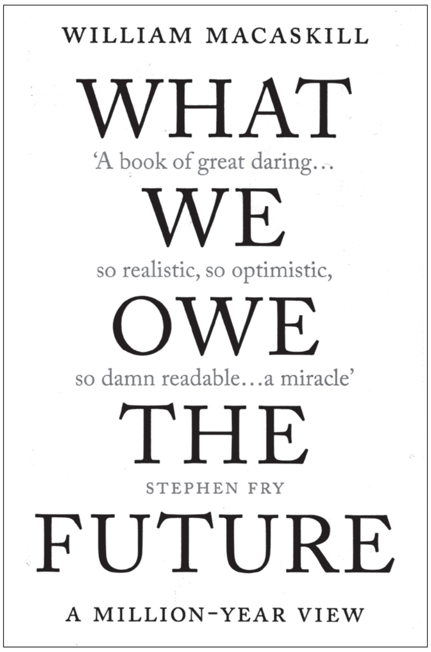
William MacAskill:
What We Owe the Future: A Million-Year View

Reviewed by Grace Clover



As a teenager living in Glasgow, the philosopher William MacAskill enjoyed urban climbing, on one occasion putting his foot through a skylight and narrowly escaping puncturing his internal organs on broken glass. At the time, he saw the likelihood of falling and dying as insignificant and thus untroubling. But now aged 35, MacAskill admits that his youthful insouciance was foolish, not because his death was *likely*, but because it 'wasn't *sufficiently unlikely*' to warrant risking such severe consequences (39). This is how MacAskill – a founding member of the Effective Altruism movement, now a researcher at the Global Priorities Institute at the University of Oxford – represents current generations in this book: as a short-sighted teenager, obliviously making decisions which will impact its long-term future. While we cannot exactly predict the likelihood or value of existential risks, he argues that they are now far too likely to remain overlooked.

MacAskill's latest work, *What We Owe the Future,* is indicative of a wider trend within the Effective Altruism movement in the last ten years, which has seen its priorities shifting away from utilitarian charitable spending on global poverty towards a greater concern with existential risks and the entrenchment of global values. His book offers a moral justification for longtermism and a framework for dealing with uncertain expected value. As in his previous book *Doing Good Better* (2015), MacAskill calls upon the reader to take a rational and disimpassioned approach to improving the world, challenging the assumptions which guide our actions, and leading us to seemingly counterintuitive but logically argued conclusions.

MacAskill calls 'longtermism', understood as an ethical theory, a "key moral priority of our time." (3) His justification for longtermism is as follows: People in the future could exist, and there could be a lot of them (9). These people should matter no less, morally, than people alive today. He writes: "I am not claiming that the interests of present and future people should always and everywhere be given equal weight. I am just claiming that future people matter significantly." (11) So long as these people live sufficiently happy lives (he does note ethical and practical problems in measuring this), it is of moral value that they are able to live. Even if the human race only exists for a fraction of the evolutionary lifespan of the average mammal (one million years), billions of people could still live in the future. This foundational thought underpins the rest of the book.

MacAskill begins by considering how we can improve the value of life in the future, theorising about how we can ensure that the future is a 'morally exploratory world', which prioritises improving wellbeing (99). He suggests that historical value changes – such as the abolition of slavery – were the contingent outcomes of one value system becoming culturally 'fitter' over time and outcompeting others, partly due to the work of activists. However, he warns that in the future such moral progress may become increasingly difficult due to a 'value lock-in' caused by the premature convergence of a global culture or by the creation of an artificial general intelligence able to implement its own values or those of a specific group. To demonstrate this, MacAskill employs a metaphor of history as molten glass, with periods during which our values are malleable, before the glass sets and they become enduringly entrenched.

MacAskill then moves onto existential risks, assigning one chapter respectively to extinction risks, civilisational collapse, and technological stagnation. MacAskill emphasises the risk posed by developments in nuclear warfare and engineered pathogens, even if they are never intentionally deployed: after all, lab leaks and nuclear false alarms occur with alarming frequency. Such catastrophes, if not causing extinction, could also drastically reduce our ability to collaborate internationally on other risks, such as climate change. Here civilisational collapse is defined a non-extinction threat through which we lose 'the ability to create most industrial and post-industrial technology'(124). He does remain optimistic, however, and notes that in the past, mankind as a whole has been remarkably resistant to catastrophes such as epidemics and global warfare. He suggests that an existential risk scenario, such as a nuclear winter, would not affect the entire globe equally, likely leaving areas such as Australasia relatively unharmed. He poses this as positive, as it would allow our species to survive, re-industrialise, and re-develop. Finally, MacAskill emphasises the risk posed by technological stagnation, arguing that as global birth rates slow, so must the rate of technological development. He sees this as detrimental to our ability to respond to existential risks. To this end, he cautions against the familiar environmentalist narrative that having children is unsustainable and instead promotes having children as a way to personally ward against civilisational collapse.

Following this, MacAskill introduces a theory of population ethics influenced by the moral philosopher Derek Parfit (1942–2017). He argues that the biological extinction of the human race would be, morally speaking, significantly worse than a non-extinction risk that killed 99.9% of the world's population, as it would prevent the lives of millions of people who might otherwise have lived in the future. He critiques the logical asymmetry of the 'intuition of neutrality', a philosophical viewpoint which sees bringing an unhappy life into the world as morally bad but bringing a happy life into the world as morally neutral (171). Instead, MacAskill argues – with some caution – that only 10% of the world's population today have below-neutral wellbeing, and thus on balance, the future will more likely be good than bad for the people living in it. He predicts that global wellbeing will increase overtime, drawing a causal relationship between increased wealth, happiness, and moral progression (assuming that we avoid value lock-in and stagnation). As such, he suggests that we have a moral obligation to ensure that future populations are able to live, and potentially grow indefinitely.

Finally, MacAskill offers practical advice about what individuals can do to implement longtermism. The arguments here are mostly familiar from his earlier writings: he suggests that the focus on personal consumption in the environmentalist movement is often misplaced and instead emphasises the good individuals can do by donating to effective charities, having children, and making well-considered career choices.

Structurally the text might have benefited if the discussion of population ethics presented in *Part IV: Assessing the End of the World* had immediately followed the moral argument for longtermism in part I, but otherwise the book's argument proceeds logically. MacAskill could have also focused more on the impact of longtermism on ecosystems and non-human animals, which remain largely overlooked. That said, the book is expansive in scope and very coherently written. As a philosopher, MacAskill is no stranger to the use of thought experiments to justify extrapolating moral positions. It is perhaps more impressive that his case studies from fields as diverse as history and zoology are so effective and evocative. The moments of personal reflection about his own life as well as his friends and colleagues also offer particularly engaging touches of warmth.

Implicit however in every part of the book – from his metaphor of humanity as a singular teenager, to his use of aggregate moral value and quality-adjusted life years – is MacAskill's treatment of humanity as an individual, rather than a collective made up of many parts with independent needs. Such a premise is foundational to his utilitarian emphasis on doing the maximum amount of good for humanity as a whole, whilst avoiding emotional assessments of individual need. This dehumanising tendency could easily alienate many readers from his conclusions. For example, most people who are concerned about climate change would agree that decarbonisation is key to our path to a sustainable future, improving the health of current people and the safety of future generations. This view is entirely coherent with the model of longtermism which MacAskill proposes here. He even describes decarbonisation as the yardstick for judgeing all longtermist action. However, many would find it absurd, or at least too abstract, that he justifies decarbonisation partly on the basis that we must leave easily accessible fossil fuels available for re-industrialisation following civilisational collapse. Under this logic, the deaths of billions in such a collapse are brushed aside, so as to emphasise the moral benefit of future population growth.

This is indicative of a more integral problem with MacAskill's work: his unwillingness to engage with the practical and emotional implications of death, or the social systems which underpin global suffering. MacAskill does note that it is a 'colossal injustice' that developing countries who contributed least to the climate crisis are likely to be most impacted by it (36), but he fails to engage with what this injustice means in practice: the intense suffering caused by drought, flooding, famine, and natural disasters, and the lack of financial resources to recover from it. Nor does he indicate any global structural changes which could even out this injustice, such as the Loss and Damage Fund agreed upon on the world climate conference in Sharm el-Sheikh in 2022. As his earlier work has shown us, MacAskill is certainly not ignorant of global suffering. But in emphasising the moral obligations we have for the future, suffering in the present appears to have lost some of its emotional weight. Regardless of what one thinks about longtermism, in an ideology framed around improving wellbeing, this seems like a contradiction.

Despite this, MacAskill offers an urgent but upbeat call to action to deal with existential risks, written in an accessible and engaging style. Though MacAskill remains deliberately cautious when drawing conclusions about the future and warns against complacency, the overriding impression left by *What We Owe the Future* is an optimism about our ability to positively impact the longterm and about the expected value of the future itself.

MacAskill, William (2022): What we owe the future: A million-year view. London: Oneworld Publications. 352 Pages. ISBN: 9780861542505 (hardback), ISBN: 9780861542512 (paperback). Price: hardback £20; paperback £10.99.