Ethics in Progress (ISSN 2084-9257). Vol. 14 (2023). No. 2,
Art. #9, pp. 130-152. DOI:10.14746/eip.2023.2.9
Creative Commons BY-SA 4.0

The Problem with Longtermism

B.V.E. Hyde

(University of Leeds, UK; University of Religions and Denominations, Iran; Durham University, UK; b.v.e.hyde@outlook.com)
ORCID: 0000-0002-1574-143X

Abstract: Moral circle expansion has been occurring faster than ever before in the last forty years, with moral agency fully extended to all humans regardless of their ethnicity, and regardless of their geographical location, as well as to animals, plants, ecosystems and even artificial intelligence. This process has made even more headway in recent years with the establishment of moral obligations towards future generations. Responsible for this development is the moral theory – and its associated movement – of longtermism, the bible of which is What We Owe the Future (London: Oneworld, 2022) by William MacAskill, whose book Doing Good Better (London: Guardian Faber, 2015) set the cornerstone of the effective altruist movement of which longtermism forms a part. With its novelty comes great excitement, but longtermism and the arguments on its behalf are not yet well thought out, suffering from various problems and entailing various uncomfortable positions on population axiology and the philosophy of history. This essay advances a number of novel criticisms of longtermism; its aim is to identify further avenues for research required by longtermists, and to establish a standard for the future development of the movement if it is to ever be widely considered as sound. Some of the issues raised here are about the arguments for the moral value of the future; the quantification of that value with the longtermist ethical calculus – or the conjunction of expected value theory with the 'significance, persistence, contingency' (SPC) framework; the moral value of making happy people; and our ability to affect the future and the fragility of history. Perhaps the most significant finding of this study is that longtermism currently constitutes a shorterm view on the longterm future, and that a properly longterm view reduces to absurdity.

Keywords: Effective altruism; longtermism; historiosophy; ethical calculus; existential risk; future; William MacAskill.

Introduction

One of the most significant advances in twentieth century moral theory was Derek Parfit's demonstration in *Reasons and Persons* (1984) that our moral intuitions lead us inexorably towards a repugnant conclusion: for any world whose citizens are perfectly equal and all living lives of immense happiness, there is a better world filled with people whose lives are barely worth living, so long as it is sufficiently populous (Parfit 1984, 388). Mankind is, therefore, better off now than before the first man fell from the Garden of Eden, because there are so many more of us now.

Philosophers have spent the last forty years trying to avoid this repugnant conclusion whilst retaining the moral intuitions that entail it. Not William MacAskill. A couple of years ago, he released a statement with over two dozen others in which he announced that avoiding the repugnant conclusion should no longer be the central goal driving population axiology, and that entailing the repugnant conclusion should not be considered an inadequacy (Zuber et al. 2021). Barely a year later, he released *What We Owe the Future* (MacAskill 2022) in which the only reason he provides for thinking this is that the premisses from which the repugnant conclusion follows are "close to indisputable" (181), apparently failing to realize that, far from being grounds to accept it, this is precisely what makes problematic the mere addition paradox (the name given to the logical reasoning that derives the repugnant conclusion from our moral intuitions). It is necessary that he takes this attitude, though, because he espouses what is effectively an applied version of the repugnant conclusion: "a civilisation that is twice as long or twice as large," argues MacAskill, "is twice as good" (MacAskill 2022, 189).

This position is part of a theory called 'longtermism' which was perhaps first brought to prominence by Toby Ord in *The Precipice* (2020) who defined it as a "moral re-orientation toward the vast future" (Ord 2020, 52). But why should you care about the future? First of all, longtermists are believers in impartiality and the equal consideration of interests – a principle utilitarian in nature – which means that, according to Peter Singer (1972), perhaps the principal utilitarian philosopher of our time, "it makes no moral difference whether the person I can help is a neighbor's child ten yards away from me or a Bengali whose name I shall never know, ten thousand miles away" (Singer 1972, 231–232). Singer also defended this in *The Life You Can Save* (2009), as did Peter Unger in *Living High and Letting Die* (1993). They argued that, because 85% of the global variation in earnings is determined by location, and because it is only by chance that you were born where you were, it is your moral responsibility to do something to try to end the poverty you know to exist in developing nations. To be a good person, they say, you must seek to do the greatest good for the greatest number, indiscriminate in whom the recipient of your charity might be.

This principle of impartiality has been instrumental in a process known as "moral circle expansion," first properly explicated by W. E. H. Lecky in his *History of European Morals* (1869) but most famously expressed by Singer in *The Expanding Circle* (1981). Integral to moral progress, this refers to the process of increasing the number and type of entities given moral consideration over time, starting with women's rights movements and later including black rights, animal rights, plant rights and even the extension of moral agency to artificial intelligence. Longtermism represents yet another expansion of the moral circle to include future people within its remit. What licenses this expansion is that, as MacAskill says, "distance in time is like distance in space" (2022, 10). According to longtermists, if we are to care about a Bengali ten thousand miles away, then we ought to care about one ten thousand years into the future.

Another reason why the future matters to MacAskill is that the future is so large and so populous (ibid. 12). This is another principle that longtermists take from utilitarianism: in their view, it is our moral obligation to enact the greatest good for the greatest number of people. One of the most (in)famous expositions of this was made by William Godwin in his *Enquiry Concerning Political Justice* (1793), who maintained that, faced with the choice of saving one of two people from a burning building, the illustrious Archbishop Fenelon of Cambray or his chambermaid, you should save the archbishop, even if his chambermaid were your wife, your mother or your benefactor, because saving the archbishop would do more good for more people (bk. 2, ch. 2). Now, this is a very objectionable idea, and one for which utilitarians received a lot of criticism these last two centuries. Pope Saint John Paul II, for example, was a personalist, which meant that he emphasized the importance of humans as persons. He thought that a danger of utilitarianism was that it tends to treat persons as objects. As he wrote in a *Letter to Families* (John Paul II 1994): "Utilitarianism is a civilization of production and of use, a civilization of things and not of persons, a civilization in which persons are used in the same way as things are used" (§13).

None of these problematics are mentioned by MacAskill, but he seems to be aware of them, because he clunkily adds to his justification of longtermism a deontological principle completely opposed to utilitarianism. He says of future generations that, "if we recognize they are real people… then we have a duty to consider how we might impact the world they inhibit" (MacAskill 2022, 19). This statement comes out of nowhere and is a rehashed version of Immanuel Kant's "formula of humanity" which he laid out in the *Groundwork of the Metaphysics of Morals* (1785): "act that you treat humanity, whether in your own person or in the person of any other, always at the same time as an end, never merely as a means" (429). It does not seem to strike MacAskill as problematic that Kant was referring to conscious persons with moral autonomy who are, crucially, alive, and not to the mere idea of possible people who do not exist but might yet still.

I. Quantifying Future Value

Just how important is the future to longtermists? They have two ways of calculating value. The first is with a "significance, persistence, contingency" (SPC) framework (MacAskill 2022, 33). This is basically a simplified version of the felicific calculus, including intensity and extent (significance), duration (persistence), and certainty (contingency) whilst excluding propinquity, fecundity, and purity. The quantification of ethics first emerged under the auspices of mathematical science in the seventeenth century. Thomas Hobbes, for example, wrote in the epistle dedicatory to *De Cive* (1651) that, "were the nature of human actions as distinctly known, as the nature of quantity in geometrical figures (…) mankind should enjoy such an immortal peace, that (…) there would hardly be left any pretence for war" (3). By the eighteenth century, ethical calculus had gotten fully underway: Francis Hutcheson, in his *Inquiry into the Original of our Ideas and Beauty and Virtue* (1725), thought that it was possible to calculate the moral importance of an agent by multiplying the benevolence of an agent by their ability. This was problematic at the time and many of its problems have persisted until now, but MacAskill does not address them.

The second method of calculating value is expected value theory, which is to measure total value as the likelihood of something occurring multiplied by how long that thing will last if it obtains. So, if there is a 20% chance of an intervention lasting a million years, its expected value is two-hundred-thousand years. This means that a longtermist would be committed to many actions with extraordinarily low chances of success as long as they are persistent enough were they to obtain. If something has a 0.0001% chance of lasting a trillion years, it follows that it is extremely valuable with an expected value of a million years. This represents the worst of all possible worlds with respect to the present, for it entails that it would be preferable to sacrifice the happiness of the present for a negligible chance of securing a trivial level of welfare, so long as that welfare would be secured for trillions of years. And thus we have the repugnant conclusion: if the future is long enough and sufficiently populous, it would be better to give them lives barely worth living than to ensure that we all live splendid lives now. Only, there is an even more repugnant aspect to longtermism; namely, we are sacrificing the whole present for the sake of a future that may never even come to pass. Theories that face the difficulty of the repugnant conclusion at least guarantee in hypothetical terms the achievement of the alternative world with a greater quantity but lower quality of welfare. Longtermism, on the other hand, is not only to sacrifice quality for quantity; it is to sacrifice it for the *mere prospect* of a more than proportional quantitative increase in welfare. That makes longtermism perhaps the most perverse population axiology to have ever been proposed.

One of its errors is the assumption that we are the only ones able to affect the whole future. If we have measures that have a 100% chance of success but only last a hundred years, the next generation can also enact those measures. However, MacAskill has no conception of accumulation, which is why his expected value theory completely neglects repeatability.

This is also reflected in the absence of a sell-by date on longtermist ethics and policy. Concern for the longterm future is as much of a moral obligation for the people of the longterm future as it is for us. We are obliged to surrender to the future our present resources that facilitate our happiness in order to facilitate the happiness of future people, who themselves will be required to forfeit those resources and their wellbeing for people future to them. The future, no matter how far away, will at some point become the present, for there is no fixed point in time that is *the* 'future'. The logical conclusion of longtermism, then, is to steal resources from all humanity throughout history and reserve them for some mystical 'future' that never comes to pass.

If this mystical future is thought of as being sufficiently large, then it might also be thought of as being more important than the present too. Longtermism is the moral view that the future is morally important; the view that the future is *more* important than the present, and that it ought therefore to be the principal moral priority of our time, is called *strong* longtermism. MacAskill makes the case for strong longtermism elsewhere (MacAskill & Greaves 2021), but not in his book. His views do not seem entirely discrete, however, and it seems difficult to separate the more radical position from the more moderate when you take its premisses to their logical conclusions; nor is anything said by MacAskill about a limit to the longtermist principles espoused that would suggest that they ought not to be taken to their logical conclusions.

II. The Intuition of Neutrality

Ignorant of these problematics, he thinks his reasons sufficient to claim that there is a "tyranny of the present over the future" that needs to be toppled (MacAskill 2022, 9); that future lives count just as much as our own (ibid. 5), or that they at least count for something (ibid. 11). However, one of the chief difficulties for longtermism is that future people do not exist yet, so MacAskill must justify why it is good to make happy people. To do so, he tackles the "intuition of neutrality" (ibid. 171) which is, in the words of Jan Narveson (1973), that "we are in favour of making people happy, but neutral about making happy people" (80). MacAskill has four arguments against this intuition – which is a little odd, because they do not seem cumulative, so one ought to have been sufficient if it were sound.

First, he assumes that our intuition is asymmetrical, such that we are indifferent about the moral value of making happy people but think that it is bad to bring a new person into existence with an unhappy life. If it is true that our intuitions are asymmetrical like this, then any argument that it is bad to bring an unhappy person into existence should work for why it is good to bring into the world a happy person (MacAskill 2022, 172). MacAskill does not give any inclination of what such an argument would look like, but quicky digresses to his second argument, which is even worse than his first.

Namely, he claims that, because it is intuitive to him that the future is better because of the existence of his happy nephews and nieces, it follows that the world is in fact better with the creation of happy people (ibid. 172). One must wonder with such an argument whether MacAskill has properly educated himself as a philosopher, for with this he seems little more than an everyday moralist, no different from every Jan Rap and his companion. Not only is this argument based on anecdotal evidence, but it begs the question and is an inductively fallacious universalization on two accounts: one, not every niece and nephew would make the future a better place – it is an open question whether it can even be said that *most* of them would; two, not everyone shares his intuition about his nieces and nephews.

His third argument is based on evidence rather than (il) logical reasoning and is in this respect better than his first two. He cites a recent psychological study that found that our intuitions about the moral value of bringing happy or unhappy people into the world are in fact symmetrical: that is, what we actually think is that it is good to create happy people and bad to create unhappy ones (Caviola et al. 2022). This means that the current evidence does not support the existence of the intuition of neutrality. Of course, this contradicts his first argument but, seeing as this one is much stronger, it is best to ignore his first. An argument from a psychological study on intuition, even if high-powered, still does not make a strong case though, not least because it is one thing us all thinking something and another matter entirely as to whether what we think is true. This runs afoul of David Hume's law, which he explicated in *A Treatise of Human Nature* (1739): that no moral statement can be inferred from non-moral ones (bk. iii, pt. i, §1).

MacAskill puts a lot more effort into his fourth argument, but that is because it requires the most mental gymnastics. It starts with the claim that present actions have drastic impacts on the future. His demonstration of this is that the chance of the spermatozoon responsible for your existence fertilizing an ovum is a mere one in two hundred million. A tiny change of timing, according to MacAskill, would therefore have resulted in someone different being born instead of you. This is what he calls the "fragility of identity" (MacAskill 2022, 173). According to him, we are "like clumsy gods" (ibid. 174), dramatically changing the course of history every second of every day. The implication is that policy today will affect the future, not by improving lives of future people that would have existed anyway, but by *creating* a new future with new people that are slightly happier. And, because it is intuitive that we have in fact improved the future, it must follow that it is good to add people with happier lives, so the intuition of neutrality is false (ibid. 176). Again, this argument begs the question, for MacAskill assumes that improving the future is morally good for us now. This is not the main problem with it, though. Where this argument is obviously invalid is in the meaning of "adding happy people," where MacAskill is guilty of equivocating: it is not at all the case that saying that it is good to improve the future is the same as saying that it is good to add happy people. Regardless of whether his little venture into reproductive biology shows that identity is as fragile as he claims it is, it has no relevance to what we *mean* when we speak about *changing* the future, and we do not mean *creating* it. This argument is best described as an inane trick. It is the kind of puzzle you give to your friends down at the pub after your first undergraduate philosophy class; you all argue together as they insist you are wrong but struggle to put into words why you are. It is not impressive at all to see a professional philosopher – one of the most influential in the world right now – touting such a mind game as if it were a philosophic proof. In fact, it seriously undermines the integrity of the whole longtermist movement that founds itself largely upon his book, and it is conceivable that such sophistry will do damage to the profession of academic philosophy too.

The fifth and final argument offered by MacAskill is the most sophisticated, but it is not his anyway: he admits by way of an endnote that he takes it from John Broome's book, *Weighing Lives* (2004). What MacAskill does do is frame it in a more accessible manner; what he does not do is try to make it any better, which is why this argument is also invalid.

Say that in world$_1$ you are not born, in world$_2$ you live in suffering and in world$_3$ you live blissfully. The intuition of neutrality says that world$_1$ is neither better nor worse than world$_2$, which means that world$_1$ is equal in value to world$_2$. Why this inference? Because Broome assumes that the comparative value relation is complete (Broome 2004, §10.1), which means that if something is neither better nor worse than something else, the two are equally as good as one another. From the intuition of neutrality it also follows that world$_1$ is just as good as world$_3$. If values are transitive, which both Broome and MacAskill assume they are, then it follows that world$_2$ is just as good as world$_3$ which, according to MacAskill, is a "contradiction" because it cannot be the case that creating a life of suffering is just as good as creating a life of bliss (MacAskill 2022, 177). Like MacAskill's other arguments, this also begs the question, for he assumes that the welfare of hypothetical additional people is not a neutral matter, and that it is therefore possible to say that adding good lives is better than adding bad ones. The argument is circular: MacAskill makes value judgements about the welfare of hypothetical additional people to justify making value judgements about the welfare of hypothetical additional people. The opposite way of framing the argument is equally invalid. If we start from the premisses that world$_3$ is better than world$_2$ and that, by virtue of the intuition of neutrality, world$_1$ is equal in value to world$_2$, then it follows that world$_3$ is better than world$_1$ as well, which disproves the intuition of neutrality. It is, of course, only by disproving the intuition of neutrality that we can say that world$_3$ is better than world$_2$ in the first place though.

MacAskill believes himself to have demonstrated that it is good to create happy people. This leads him to the conclusion that having children is good (MacAskill 2022, 187) and that we ought to ensure that civilization lasts as long as possible and is as big as possible (ibid. 188). In other words, the bigger the future, the better the future, which is why "the early extinction of the human race would be a truly enormous tragedy" (ibid. 189).

III. The Longtermist Image of History

Assuming, then, despite wanting for a single sound argument to its effect, that it is good to make happy people, how do we go about doing so? Well, we first need to be able to actually affect the future. Perhaps MacAskill does not realize it, but he is committed to a certain philosophic position on history and causation. Namely, longtermism assumes that history is not deterministic. G. W. F. Hegel's *Phenomenology of Spirit* (1807) is perhaps the most famous example of a teleological theory of history in which the march of history represents mankind's ascent to consciousness of the world, or the spiritualization of nature, such that the future is destined to be more conscious than the past – a view that has more recently been taken up by Francis Fukuyama in *The End of History and the Last Man* (1992). A slightly less deterministic example is Thomas Carlyle who, in his book *On Heroes, Hero-Worship, & the Heroic in History* (1841), said that "the History of the world is but the Biography of great men" who drive history in pursuit of greatness (ch. 1). Karl Marx flipped the operative great man theory on its head in *The German Ideology* (1932) but, nevertheless, he still thought that history was teleological. Instead of individuals who steer history, he propounded a contextual historiosophy known as historical materialism in which the course of history is determined by a series of class struggles which terminate with communism at the end of history. Likewise, any religious view of history would struggle to avoid some degree of determinism, not only through divine providence but because history must inevitably end with the final judgement in which the sheep are divided from the goats (Matthew xxv. 31–46) and the messiah ushers forth the Kingdom of God.

Such grand narratives are not so popular today, but it is not at all obvious that historical causation is as fragile or as chaotic as MacAskill sometimes makes it out to be. For instance, it is questionable to what extent it is really true that your identity is so fragile that a tiny change of timing several decades ago would have prevented your birth. What MacAskill is inadvertently committing himself to is chaos theory and an extremely sensitive dependence on initial conditions. Alan Turing (1950) was somebody who claimed that the displacement of even a single electron by a billionth of a centimetre at one moment might make the difference between life or death for a man a year later. A more expressive way of putting it, and one of the most quoted – even if it is a little inaccurate – is Edward Norton Lorenz's butterfly effect: "Does the flap of a butterfly's wings in Brazil set off a tornado in Texas?" MacAskill is obliged to say that it does, but this is a difficult position to defend and one that he makes no attempt to, unaware that he has even cornered himself like this.

Even if he cannot justify sensitive dependence on initial conditions, MacAskill does put forward a fair case for our ability to affect the future in some respects. One of his better examples is the anthropogenic extinction of species (MacAskill 2022, 30), which would suggest that we can have an irreversible influence on the future. More important for his case, though, is that changes in values are to demonstrate their contingency and thus our power over the future (ibid. 62). The abolition of slavery is MacAskill's proof by example. Slavery, he thinks, was not abolished due to economic changes, as many historians have maintained, but a change in values (ibid. 70). What caused this change in values, he does not really say. He pins it on individuals and their activism, but does not have a sophisticated historiographical interpretation of how this led to social change or why this social change happened in the nineteenth century rather than earlier. This is, ultimately, because he does not have any explanation for why moral progress occurs, which is a discussion omitted entirely from MacAskill's argument, which means that he cannot exclude the possibility that values change as part of a fully determined march of moral progress throughout history. In such a case, changing values does not demonstrate our ability to change the future at all.

It is not necessary for the validity of MacAskill's argument that history is as fragile as he sometimes suggests it is. In the spirit of the principle of charitable interpretation, we might grant that only significant changes to the present will result in any real change in the future, and we might also allow him that we can in fact change the future, even if he has not demonstrated this, at least because it is intuitive. This still raises numerous questions.

Firstly, is anything else contingent other than moral values? We can all agree that morality is very volatile, especially at present, and that it is therefore perhaps contingent and can be determined or redetermined by us, but what about political values, or scientific values? What about things other than values, like nature? It seems like we can destroy things like species with relative ease, but what about recovering endangered species, or affecting the atmosphere? Can all these things be changed the way we change values? Are we really the clumsy gods that MacAskill claims we are, where absolutely everything changes at our whim?

Secondly, will those things of ostensible concern to the future actually have any significant impact? Would, for example, the world be *so* different now if the dodo had not been hunted to extinction? Would raising the global temperature by a degree or the sea level by a few inches two thousand years ago have prevented the Romans from invading Carthage? Would it have prevented the development of civilisation around the Mediterranean? When put like that, it seems farfetched. Even if we let the sea level rise by seventy metres and the world warm by several degrees, dikes will be built the same way that land is reclaimed from the sea in the Netherlands today and, although agricultural practices might change, it is unlikely we will be completely without food. MacAskill himself points out that humanity would even survive nuclear winter (MacAskill 2022, 131). This is not to say that we should not care about these things: on the contrary, if we are forced to flee the earth and terraform Mars, then perhaps they will have an enormous impact. However, there is certainly an open question about the relative importance of many things to human history, and it has often been the case that what people have thought important at one time has proven to be largely unimportant at a later date. MacAskill assumes that our present actions will have significant future effects, but he has done little to explain why this is the case.

Thirdly, what about a properly longtermist view? MacAskill looks at abolitionism as an example of how we can change values, but these are recent moral developments. What about developments over many millennia? After all, longtermism is supposed to apply over millennia, even millions of years. Can a case be made that our civilization is today defined by the first men many thousands of years ago? It seems unlikely. Can *any* influence of our forty-thousand-year-old ancestors be seen today? The further into the future you look, the more questionable it is to what extent you can actually do anything about it. For sure, you have time on your hands, but a lot can happen in a million years. It is not immediately obvious that even fairly significant events now will have an effect on the future, especially one so many millions of years long.

MacAskill does attempt to answer these questions. Values, he thinks, can persist for extremely long periods of time through “value lock-in” (MacAskill 2022, 78), and he gives the examples of Confucian influences on the Orient today and Christian influences on the modern Occident, though he does not complete the stories he tells to explain how such values are still instantiated in these contemporary cultures. This is a common claim though, and there is no reason to disbelieve it. It is well-documented in other works (i.a. Tu et al. 1992; Perkins 2004; Woods Jr. 2012; Zhang 2016). Paul Tillich captured the essence of these influences best in *The Protestant Era* (1948) when he said that “religion is the substance of culture and culture the form of religion” (57).

According to MacAskill, the permanence of values is determined by an “early plasticity, later rigidity” cycle (MacAskill 2022, 42). History is like glass that is sometimes hot and sometimes cold. When it is hot, it can be reshaped, but the colder it gets the harder it becomes. This is a fairly accurate estimation with respect to historical change that many historians will likely be inclined to agree with. There are, clearly, periods of change and periods of stagnation in history: compare, for example, the Renaissance and the Enlightenment with the Dark Ages, or look at Thomas Kuhn’s observation in *The Structure of Scientific Revolutions* (1962) about how scientific paradigms are toppled by scientific revolutions.

Right now we are at the crossroad of history or, as Derek Parfit wrote in his book *On What Matters* (2011), we “live during the hinge of history” (vol. 2, 611). The present age is one of plasticity, but we are warned that a period of rigidity is on the horizon. A lot can happen in hundreds or thousands of years, but MacAskill thinks that not a lot will – at least when it comes to values. A value lock-in is coming. What will cause it, he says, is artificial intelligence: because it is immortal and has the potential to cause rapid technological progress, whatever values it holds, or whatever values are instilled within it, could last a really long time – maybe even forever (MacAskill 2022, 83). This means that our values could define the future, which is why changing them for the better is one of the most important longtermist tasks (ibid. 52).

This time around, artificial intelligence will cause the future to set its course, but what kinds of things have systematically determined plasticity and rigidity in the past, and will they continue to do so? MacAskill does not say, which is a considerable oversight and yet another instance of his lack of awareness with respect to the sociological and historiosophical ramifications of his claims. The impression he gives is that this cycle is just an intrinsic part of the nature of the passage of time. This would suggest that MacAskill is a social cycle theorist, the central idea of which is that societies naturally move through repeating cycles of growth, stagnation, decline and regeneration. However, there are lots of theories that would suggest that there are identifiable causes of change and stagnation. Sociocultural evolutionists, for instance, posit that societies undergo stadial evolution driven by technology, population, environment, and social organization. The theory draws on ideas from both biology and anthropology and is based on the premiss that human societies are subject to similar selective pressures and evolutionary processes as biological organisms. Another theory, developed by Thomas Malthus in his *Essay on the Principle of Population* (1798), is that societies face recurring cycles of growth and stagnation as they struggle to balance population size with available resources. Another explanation is that power determines plasticity and rigidity. When incumbent values and institutions are powerful, they are resistant to change; what accounts for an end to sociocultural stasis is a transfer of power. In his *Treaty of General Sociology* (1916), Vilfredo Pareto divided the social elite into cunning foxes and violent lions. In his view, power is constantly passed between the foxes and the lions. This image is reminiscent of the dialectical, historical materialist and great man theories of history. Whether it be values, classes or great men, there are heroes and villains in history, the course of which is determined by the struggle between them. When there is a hegemony of the one, there is rigidity.

It is a little odd that MacAskill, rather than focussing on creating a democratic world in which nobody coerces another or forces their values upon them, instead chooses to take advantage of moral and institutional hegemony. Regardless of whether the values he is trying to impose are good or not, that is a somewhat insidious approach to take. What is more, as he rightly points out – in his most recent book (MacAskill 2022, 187), his book on *Normative Uncertainty* (2020) with Toby Ord and Krister Bykvist, and his doctoral thesis on the same topic – we currently operate under moral uncertainty, meaning that he cannot even be sure that the values he is trying to stick the future with are even the right ones. Whether it is the power of cyclical historical change, sociocultural evolutionary factors, or history's actors, to take advantage of that power is to become the oppressor.

MacAskill does claim, however, that it would be better to avoid value lock-in (MacAskill 2022, 88). Ideally, he would have a "long reflection" (ibid. 98): a long period of stability where humanity can reflect on the idea of the good to work out what a flourishing society would look like. What he is striving for is a "morally exploratory world" in which better morals win over time such that we converge on the best society (ibid. 99). For this to happen, value lock-in must be avoided, which means that there are a few things we need to do. One: we must prioritize the prevention of value lock-in, even at the expense of delaying advancement such as space exploration or development of artificial intelligence. Two: we must be politically experimental and ensure that our society is culturally and intellectually diverse to avoid premature convergence. Three: we have to somehow ensure that cultural evolution results in moral evolution. What we end up with is a "lock-in paradox" (ibid. 101): we need to lock-in some institutions and values to prevent a more thoroughgoing lock-in of values. That is, we must become the ideological oppressors, else there might be worse ideological suppression. This seems to run up against his precept that the ends do not justify the means (ibid. 240).

As much of a drama that MacAskill makes about value lock-in, whether or not our values will determine the longterm future does not make any tangible difference to the way we conduct ethics now. We still ought to live our best lives, we still ought to better society and its values and, when it comes to us philosophers, we still ought to seek the truth of the good. What benefit is it to think that there is a deadline for all this? Regardless of any responsibility for the future, a morally exploratory world is something sought after anyway. Nobody wants a global moral hegemony. All MacAskill has done is try to instil some egotistical sense of righteousness and self-importance into our everyday moralizing and the ethical theorizing of professional philosophers like MacAskill himself, but no change will be effected because of this. Is the pursuit of moral truth not enough for him? Must his moral musings define the future of mankind? Is his work *so* important? It does not change the nature of the work though. We will still pursue the same moral ideals as always; only, now many (i.e., longtermists) will do so on a high horse feeling all mighty about themselves that their moralizing is all the more important. The future is in *their* hands. They are the clumsy gods that future civilizations ought to worship.

What is more, that our values last a long time does not mean that we need to have a longterm view, even if we are morally obliged to the future, even if we are its artificers. It is enough to attend to the present; worry that you are acting well now. A moral exemplar for many does not need to see to the wellbeing of his followers. If he is a good man and they take him as example, they too will be good men, and we will have many flourishing generations of good men. It does not make a man any less charitable because those influenced by his morals are not recipient of his charity; nor does it make him any less good because his disciples are not the objects of his goodness; nor even is he for the worse because he is good without thought for those who follow him. That you are responsible for the moral values of the future does not mean that you need to concern yourself with it.

IV. Existential Risk Prevention

Avoiding a lock-in of malign values will help to ensure that the future is bright, but the future can only be good if there is one in the first place. To the extent that it is our moral obligation to ensure that the future is as big and as long as possible, we have a responsibility to avoid extinction, societal collapse, or technological stagnation too. These are existential risks, which Nick Bostrom (2013), one of the main figures to draw attention to them, defines as a threat to the premature extinction of intelligent life on earth or the permanent and drastic destruction of its potential for desirable future development.

The principal risks of extinction are, according to MacAskill (2022), engineered pathogens (107) and war between great powers (114). Traditionally, futurologists have been very troubled by the prospect of nuclear warfare. It was due to the dawn of the atomic age with the bombings of Hiroshima and Nagasaki that the *Bulletin of the Atomic Scientists* was established in 1947, maintaining the Doomsday Clock that represents the likelihood of impending catastrophe. In the last decade, however, concerns have centred around artificial intelligence due to rapid developments in the technology. The past couple of years have seen attention return to the threat of nuclear warfare, though, with the Russo-Ukrainian war. The Doomsday Clock was turned to ninety seconds before midnight in January of 2023 – the nearest it has ever been – with the reason cited being the war. MacAskill wrote his book during the war, so it is expected that he would rate the existential risk of nuclear warfare quite highly.

Many futurological researchers are most concerned by existential risk from artificial general intelligence, where humans could be replaced as the dominant lifeform on earth were machine brains to surpass human brains and become superintelligent – a view expressed by, most notably, Nick Bostrom, in *Superintelligence* (2014), James Barrat, in *Our Final Invention* (2013), and Stuart J. Russell, in *Human Compatible* (2019). Some are skeptical of this alarmism, like Michio Kaku who, in *Physics of the Future* (2011), said that he believed we will find intelligent robots benevolent and friendly. In a sense, MacAskill belongs to both groups, for he believes that artificial intelligence still represents intelligent life with moral value, so even its destruction of humanity would not be a crisis so long as the artificial civilization that advances into the future is not morally bankrupt (MacAskill 2022, 87).

As well as extinction, societal collapse could cut the future short. The main risks, in MacAskill's (2022) esteem, are climate change (134) and fossil fuel depletion (138) because, if the environment is inhospitable, or if civilization lacks fossil fuels, it is unlikely that humanity will be able to recover from societal collapse. In itself, the collapse of civilization would not spell the end of civilization: MacAskill is optimistic about humanity's ability to recover and redevelop, which is in marked contrast to Brandon Carter's (1983) controversial Doomsday Argument, which hypothesizes that humanity has a 95% chance of extinction within the next ten thousand years, with J. Richard Gott III (1993) estimating this within the next 7.8 million years. Small examples of civilizational resilience have been seen throughout history, such as the recovery after the Sack of Rome in 1527. A more dramatic, and more speculative, case would be the recovery of civilization after an alleged comet strike that completely wiped out an advanced civilization that supposedly existed during the Younger Dryas around twelve thousand years ago – a theory advanced by Graham Hancock in *Magicians of the Gods* (2015). This theory, however, is not accepted by mainstream science (Shermer 2017). If it were, it would make great evidence for MacAskill's optimism in human resilience. In any case, such recovery is contingent upon an agricultural revolution which requires a hospitable environment and an industrial revolution which, at least at the beginning, requires fossil fuels.

Both extinction and societal collapse, MacAskill thinks, are quite avoidable. Technological stagnation might not be. There is less that can be done to prevent it, and we are reliant upon the chances of either a technological breakthrough or a population boom which might lead to technological advancement – by virtue of there being more scientists – in the absence of which stagnation is likely (MacAskill 2022, 156). If we want humanity to survive, we need to make the chance of catastrophe as small as possible, then keep it small. However, if we enter a period of technological stagnation, these risks might stay at their present levels. Their eventual likelihood being compounded over the millennia, every century rolling another die – if Toby Ord (2020, 167) is right to estimate the chance of extinction as one in six in the next century – catastrophe might become all but inevitable.

V. Predicting Future Welfare

Despite these concerns, MacAskill is quite confident that the future will be good (MacAskill 2022, 193), in part because the world is good today. He raises three theories of wellbeing (ibid. 195). The first is preference satisfaction, according to which your life is good if your preferences are fulfilled. The second is the objective list view; that there are objective goods like friendship that make life good. And the third is hedonism which, according to him, is when positive experiences make your life better and negative experiences make it worse. This interpretation of hedonism diverges from the traditional one in which pleasure specifically, not experience, is the determinant of wellbeing, the term itself deriving from the Greek term for pleasure (ἡδονή). As Jeremy Bentham declared at the beginning of his *Introduction to the Principles of Morals and Legislation* (1789): "Nature has placed mankind under the governance of two sovereign masters, *pain* and *pleasure*. It is for them alone to point out what we ought to do, as well as to determine what we shall do" (1).

There are three psychological methods of measuring wellbeing as it is defined by these three theories: surveys to measure life satisfaction; surveys that simply ask people if they are happy; and asking people at random times how they feel in that moment (MacAskill 2022, 195–199). All three are problematic in some way or another, so MacAskill commissioned psychologists to run a survey asking people in India and in America various questions about the quality of their life (ibid. 199). What he found was that positive answers were much more common than negative ones. He concluded that, although around 10% of the global population have lives below neutral wellbeing, most people have positive lives (ibid. 201). The world is, therefore, good.

Moreover, the world is getting better. Richard A. Easterlin (1974) published a very famous study in which he showed that there is a paradox with respect to economic growth and human happiness: although higher income is correlated with higher happiness within and across countries at a point in time, people and countries do not get happier as they get richer over time. However, Easterlin (2022) himself has said that he concluded from the fact that we could not at the time show that countries got happier as they got richer that there was no relationship between absolute income and happiness, and that happiness was instead determined by income relative to one's peers. His findings were first published when data about levels of happiness was much sparser than today, and it has since been revealed that the Easterlin Paradox does not exist. More recent work with better data strongly supports the hypothesis that countries get happier as they get richer (Stevenson & Wolfers 2008). This is not just a correlation: there is evidence to suggest that there is a causal relationship between wealth and wellbeing. Contrary to the common belief that lottery winners are unhappy, originating with the psychologist Philip Brickman and his colleagues (1978), Andrew Oswald and Rainer Winkelmann (2019) have shown that winning the lottery does increase one's happiness. As the world gets richer, it gets happier. If this trend continues, we can expect the future to be even happier. People are naturally pessimistic though, with only 10% of people thinking that the world is improving. What accounts for this, MacAskill says, is the news, because "if it bleeds, it leads" (2022, 206). Despite this pessimism, the world is actually getting happier.

What might throw off this upwards trend is that, although the evidence suggests that we are happier than the preindustrial agriculturalists, hunter gatherers were very happy. They were so much happier than the agriculturalists that Jared Diamond (1987) calls the agricultural revolution "the worst mistake in the history of the human race". This is a very common view expressed by Marshall Sahlins, in *Stone Age Economics* (1972), and Yuval Noah Harari, in *Sapiens* (2015), among others. There has, nevertheless, been a clear upward trend in wellbeing since the industrial revolution, and MacAskill assumes that this trend will continue.

This estimation, however, ignores some of the predictions made by several recent studies. Matthew O'Lemmon (2022), for instance, describes the third industrial revolution, or the digital revolution, as "the worst mistake 2.0" and expresses concerns about what Ronald Wright calls the "progress trap" in *A Short History of Progress* (2004), where material innovation leads to uncertainty and a litany of problems which might lead to stagnation and possible collapse. Jared Diamond thought that the agricultural revolution had three drawbacks from which we have never recovered in twelve thousand years. These were class division, adverse effects on health, and the concentration of power in the hands of the few. There is apprehension that the digital revolution is producing a similar impact but at a far more rapid rate given the advances in artificial intelligence. Such worries have been voiced in some form or another for many years, with early concerns being expressed by Martin Heidegger in his *Lectures and Essays* (1954) and by Irving John Good (1966). MacAskill neither mentions nor addresses such concerns, however.

Another possible source of error in his estimate would be the welfare of non-human animals. MacAskill assumes that human happiness will increase as time goes on but acknowledges that this might not be the case for animals. At least since Peter Singer's *Animal Liberation* (1975), speciesism has been on the decline and animals have been increasingly included within the moral circle. Animal exploitation, however, is as prevalent as ever in the animal industrial complex, a concept coined by Barbara Noske who wrote in *Humans and Other Animals* (1989) that animals "have become reduced to mere appendages of computers and machines" (Noske 1989, 20). "Putting this all together," says MacAskill, "it seems hard to resist the conclusion that, when a factory-farmed chicken, pig, or fish dies, that's the best things that's happened to them" (MacAskill 2022, 209). Wild animals, however, are no better off, suffering from disease, injury, parasitism, starvation, dehydration, harsh weather, natural disasters, psychological distress, and being hunted and killed by other animals. Parasitism, especially, has been considered so abhorrent that it has been called 'the great evil of life' by Alexander Skutch (1948). Nature is, as Alfred Tennyson said in his poem *In Memoriam* (1850), "red in tooth and claw" (c. 56). The suffering of wild animals has often thought sufficient to preclude the possibility of there existing a benevolent god. William L. Rowe (1979), for example, in a much-discussed article, provides a poignant exemplum of a fawn, burned in a forest fire, suffering for days before its death. The problem of predation provides a real challenge for those who want to say that the world is good, and it is fair to assume that the world is not good for animals – not wild ones, and especially not farmed ones.

But how to calculate this suffering to weigh it against human happiness? MacAskill considers determining moral value by neuron count (MacAskill 2022, 210), which is fascinating. Ethical calculus is really at the heart of the longtermist project, but MacAskill does not properly consider the real difficulties with this, of which there are many. The most obvious problems are the ostensible subjectivity and contextual relativity of moral value, resulting in deep-seated moral disagreement. This poses such a challenge to the establishment of a universal measure of moral value that J. L. Mackie believed that the most rational inference from our inability to agree on moral judgements is that they are all false – a position called error theory which he advanced in *Ethics: Inventing Right and Wrong* (Mackie 1977). Another prominent difficulty for moral quantification is that moral values seem to be incommensurable, which was raised by Isiah Berlin in the third of his *Four Essays on Liberty* (1969). This may be because they are qualitatively distinct, arising from different moral frameworks or systems, or because they reflect fundamentally different aspects of human experience. For example, how does one compare liberty and equality? Ought individual freedom be prioritized even at the expense of equality? Or, should we prioritize equality over freedom? Because there is no agreement on such issues, it is particularly challenging to create a unified measure of moral value. In the end, it is unlikely that many will accept using neuron counts as proxies for moral weight. It is unclear whether MacAskill accepts this approach either, and he makes no clear comment on whether the combined welfare of humans and animals is positive or negative. Seeing as he thinks that the world is currently good, and that the future will be better, the assumption would be that MacAskill does not think that animal suffering outweighs human happiness.

When estimating future wellbeing, MacAskill is clearly optimistic, but he believes that there is good reason to be. Considering the possibility of either a utopia or a dystopia, whether we are optimistic or pessimistic depends on two things: the relative value of these worlds, and how likely we are to realize them (MacAskill 2022, 216).

Their relative value gives grounds for pessimism: the better world is not nearly good enough to outweigh how bad the worse world is. MacAskill points out that it the view of various positions in moral philosophy that things bad should be more heavily weighted than things good. Furthermore, MacAskill takes a "critical level view" of population axiology, which means that it is good to create a good life but only if it is *sufficiently* good; that is, above a certain "critical level" of wellbeing (ibid. 185). This means that it is bad to create lives that are good but with only low positive wellbeing and, thus, the future needs to have *considerably* more good than bad for it to be good on the whole (ibid. 217).

With respect to the likelihood of the realization of either a utopia or a dystopia, however, there are good grounds for optimism because a utopian future is much more likely. The reason for this is that "people sometimes produce good things just because the things are good, but people rarely produce bad things just because they are bad" (MacAskill 2022, 218), and people are generally trying to promote good (ibid. 219). Most evils, MacAskill thinks, though he does not say explicitly, are necessary lesser evils. This is an old principle that extends back to at least the Greeks but was most popular with Enlightenment theologians. Benedictus de Spinoza put it most emphatically in his *Ethics* (1677): "Of two things which are good, we shall follow the greater good, and of two evils, follow the less" (pt. iv, prop. 65). If this principle obtained, it would probably make the world good, at least in intention. It is improbable that it does, though. MacAskill has a naïve view that fails to take notice of the sheer abundance of evil in the world. He also neglects the depth of evil in the existence of what Marilyn McCord Adams and Stewart Sutherland (1989) call "horrendous evils", which are evils so grave that they provide prima facie evidence that life is not worth living and that the world is evil.

In Fyodor Dostoyevsky's *Brothers Karamazov* (1880), Ivan Karamazov tells a number of haunting tales to his brother Alyosha. He tells him about the crimes committed by Turks and Circassians in all parts of Bulgaria through fear of a general rising of the Slavs; about a murderer whose happiest day was when he was led to the guillotine; about a young girl who was subjected to every possible torture; and about the death of an innocent child, torn to pieces by dogs in front of his mother's eyes. Ivan then turns to Alyosha and says of the world that "it's not worth the tears of that one tortured child who beat itself on the breast with its little fist and prayed in its stinking outhouse, with its unexpiated tears." "If the sufferings of children go to swell the sum of sufferings which was necessary to pay for truth," he continues, "then I protest that the truth is not worth such a price." "Too high a price is asked for harmony," he says: "it's beyond our means to pay so much to enter on it" (bk. v, Chap. iv). One might wonder what he would think about longtermism.

In his naivety, MacAskill thinks that the greater likelihood of utopia outweighs the greater relative value of dystopia (MacAskill 2022, 220), though he does not explain his reasons for thinking so. The future will, therefore, or so he thinks, probably be good.

VI. Ensuring Future Welfare

There are some things we need to do to ensure that it ends up good, though: we must learn, increase our options, and do good (MacAskill 2022, 226). These have relevance for humanity at large but also for individuals, such as with respect to career choice (ibid. 235). To choose the biggest priorities, you take the "significance, persistence, contingency" (SPC) framework from before (ibid. 33), which measures importance, and add tractability and neglectedness to the mix to get an "importance, tractability, neglectedness" (ITN) framework (ibid. 256). Problems ought to be tractable and neglected as well as important to justify our attention, because important but intractable and popular issues are unlikely to bear fruitful results from an extra person studying them and donating to tackling them, whereas neglected and tractable ones are likely to give us much greater results in these respects, even if they are relatively less important (ibid. 230). What does the application of this give us? Research into artificial intelligence. Like probably all longtermists, MacAskill makes a big deal out of the existential threat of unaligned artificial general intelligence. In his opinion, research into it is not only important but neglected and tractable too (ibid. 231).

Moreover, he argues that philanthropy is essential to ensuring that the future is good: *The Most Good You Can Do* (2015) by Singer and *Doing Good Better* (2015) by MacAskill are the two bibles of effective altruism – a new movement that argues for a "scientific approach" to philanthropy in order to do the most good. This involves earning more to give more, then living modestly to give even more and, lastly, choosing the best causes and organizations that will do the most good with your money.

One of the real problems that effective altruists have faced is whether it is still most effective and, further, whether it is even morally permissible, to earn to give through jobs that cause harm, like working for an investment bank involved in economic colonization of third-world countries, or that invests in environmentally destructive practices. Another concern with such lucrative careers is that they perpetuate an unjust system (Todd et al. 2012). A firm consequentialist, Peter Singer dismissed these worries, arguing that the negative impact of an unethical job on your ethical giving is small because, "if you do not take the position offered by the investment bank, someone else will" (Singer 2015, 52). MacAskill used to have exactly the same attitude: just under a decade ago, he wrote an article in which he too argued that the positive impact of your donations far outweighs the negative impact of your means to acquiring the money to donate because, if you do not do the job, someone else just as qualified will (MacAskill 2014).

These ethical objections, Singer thought, "will come to be seen as typical of the grumblings of an older generation that does not really understand what the next generation is doing" (Singer 2015, 53). However, 80,000 Hours changed tack in 2017, recommending the avoidance of careers that do significant direct harm, even if it seems like the negative consequences could be outweighed by donations, because the harms from such careers may be hidden or otherwise hard to measure. Likewise, MacAskill has now admitted that he was wrong about earning to give from harmful careers: he no longer thinks the ends justify the means (2022, 240) and instead thinks that you should be ambitious but limit the risk of doing harm; that is, "target upsides but limit downsides" (ibid. 237). This reflects a considerable turn away from effective altruism's purely consequentialist thinking – a turn which can be seen in MacAskill's justification of longtermism, which sloppily combines principles utilitarian and deontological (cf. ibid. ch. 1).

VII. The Absurdity of Longtermism

MacAskill is still looking at the future from an immediate perspective: he has a shorterm view on the longterm future, not a longterm view at all. The million-year view that MacAskill claims to take is both farcical and difficult to wrap your head around. Here is what a really longterm perspective would look like: Say you have a savings account, like most savings accounts right now, with a pretty a bad interest rate – 3% for instance. Beguiled by MacAskill's appeals to emotion, you decide that you need to think of the children; that is, the children a million years away. So you put exactly $1 into this account which, you tell the bank, is to be withdrawn in a million years to solve all the future's problems. You then die peacefully, knowing that you have singlehandedly saved the longterm future. A hundred years after your death, there will not be even $20 in the account, but a thousand years after your death, your dollar has turned into almost $7 trillion. For measure, this is about 40% larger than Japan's current gross domestic product. Ten thousand years after your death, you will have bestowed upon the future about $2.35e+128, which is roughly 10^{116} times the size of the United Nation's estimate of the world's current gross domestic product at around $85 trillion. In a hundred thousand years, your donation to the future will be worth around $5.27e+1283, and after the full million years it is so large that it can only be approximately calculated as $$10^{12,837}$ which is a number so large that it is almost impossible to compare it to anything. The number of atoms in the known universe, for example, is about 10^{80}; if you were to buy worlds at $85 trillion a piece, you would still have trillions of times more worlds than the number of atoms in the universe; if you were to reduce each world to the size of an atom, you could buy more than eight hundred and fifty universes. If, rather than $1, you were to match the largest philanthropic donation ever of around $100 billion by Jamsetji Tata, your donation would be worth $$10^{12,849}$ after a million years, which is not a lot bigger than $$10^{12,837}$ really.

Of course, long before you reach this point, the global economic system would have collapsed because of your donation. This is not meant as a practical suggestion but is supposed to reduce to absurdity the idea of a million-year view, and it really makes a mockery of the whole longtermist motivation of the effective altruism movement. The future is simply too large to think about in any detail. MacAskill does not realize just how big a million years is. It is unfathomable. Everything turns to dust in that kind of time: even a hundred billion dollars ends up indistinguishable from a single dollar. Utilitarianism was criticised for being demanding in requiring impartiality, but longtermism takes those demands and puts them on steroids. The idea that we *can* even be morally concerned about what is a million years away, yet again obliged to do something about it, is utter folly. And, of course, Immanuel Kant tells us that "ought implies can" (1781, 548). The idea of longtermism is itself absurd, so it is not a surprise after all that MacAskill is incapable of justifying it.

References

Adams M. McCord & S. Sutherland 1989. "Horrendous Evils and the Goodness of God," *Proceedings of the Aristotelian Society Supplementary Volume* 63:297–323.

Barrat J. 2013. *Our Final Invention*. New York: Thomas Dunne Books.

Bentham J. 1789. *Introduction to the Principles of Morals and Legislation*. London: T. Payne and Son.

Berlin I. 1969. *Four Essays on Liberty*. Oxford: Oxford University Press.

Bostrom N. 2013. "Existential Risk Prevention as Global Priority," *Global Policy* 4(1):15–31. https://doi.org/10.1111/1758-5899.12002

Bostrom N. 2014. *Superintelligence*. Oxford: Oxford University Press.

Brickman P. 1978. "Lottery Winners and Accident Victims: Is Happiness Relative?" *Journal of Personality and Social Psychology* 36(8):917–927. https://doi.org/10.1037/0022-3514.36.8.917

Broome J. 2004. *Weighing Lives*. Oxford: Oxford University Press.

Carlyle T. 1841. *On Heroes, Hero-Worship, & the Heroic in History*. London: James Fraser.

Carter B. 1983. "The Anthropic Principle and its Implications for Biological Evolution," *Philosophical Transactions of the Royal Society A* 310(1512):347–363.

Caviola L., D. Althaus, A. L. Mogensen, & G. P. Goodwin. 2022. "Population Ethical Intuitions," *Cognition* 218, art. 104941. https://doi.org/10.1016/j.cognition.2021.104941

Diamond J. 1987. "The Worst Mistake in the History of the Human Race," *Discover Magazine*, May 1987, pp. 95–98.

Dostoyevsky F. 1880. *Brothers Karamazov*. St. Petersburg: A. F. Marks.

Easterlin R. A. 1974. "Does Economic Growth Improve the Human Lot? Some Empirical Evidence," in P. A. David & M. W. Reder (Eds.), *Nations and Households in Economic Growth*. New York: Academic Press.

Easterlin R. A. & K. J. O'Connor. 2022. "The Easterlin Paradox," in K. F. Zimmermann (Ed.), *Handbook of Labor, Human Resources and Population Economics*. Cham: Springer.

Fukuyama F. 1992. *The End of History and the Last Man*. New York: Free Press.

Godwin W. 1793. *Enquiry Concerning Political Justice*. London: G.G.J. and J. Robinson.

Good I. J. 1966. "Speculations Concerning the First Ultraintelligent Machine," *Advances in Computers* 6:31–88. https://doi.org/10.1016/S0065-2458(08)60418-0

Gott III, J. Richard 1993. "Implications of the Copernican Principle for Our Future Prospects," *Nature* 363(6427):315–319. https://doi.org/10.1038/363315a0

Harari Yuval N. 2015. *Sapiens*. New York: Harper.

Hancock G. 2015. *Magicians of the Gods*. New York: Thomas Dunne Books.

Hegel G. W. F. 1807. *Phenomenology of Spirit*. Bamberg – Würzburg: Joseph Anton Goebhardt.

Heidegger M. 1954. *Lectures and Essays*. Pfullingen: Günther Neske.

Hobbes Th. 1651. *De Cive*. London: R. Royston.

Hume D. 1739. *A Treatise of Human Nature*. London: John Noon.

Hutcheson F. 1725. *Inquiry into the Original of our Ideas and Beauty and Virtue*. London: J. & J. Knapton and Co.

Kaku M. 2011. *Physics of the Future*. New York: Doubleday.

Kant I. 1781. *Critique of Pure Reason*. Riga: Johann Friedrich Hartknoch.

Kant I. 1785. *Groundwork of the Metaphysics of Morals*. Riga: Johann Friedrich Hartknoch.

Kuhn T. 1962. *The Structure of Scientific Revolutions*. Chicago: University of Chicago Press.

Lecky W. E. H. 1869. *History of European Morals from Augustus to Charlemagne*. London: Longmans, Green, and Co.

MacAskill W. 2014. "Replaceability, Career Choice, and Making a Difference," *Ethical Theory and Moral Practice* 17(2):269–283. https://doi.org/10.1007/s10677-013-9433-4

MacAskill W. 2015. *Doing Good Better*. London: Guardian Faber.

MacAskill W. 2022. *What We Owe the Future*. London: Oneworld.

MacAskill W. & H. Greaves 2021. "The Case for Strong Longtermism," *Global Priorities Institute Working Papers Series* 5-2021.

MacAskill W., T. Ord, & K. Bykvist. 2020. *Normative Uncertainty*. Oxford: Oxford University Press.

Mackie J. L. 1977. *Ethics: Inventing Right and Wrong*. New York: Penguin Books.

Malthus T. 1798. *Essay on the Principle of Population*. London: J. Johnson.

Marx K. 1932. *The German Ideology*. Moscow: Marx-Engels Institute.

Narveson J. 1973. "Moral Problems of Population," *Monist* 57(1):62–86.

Noske B. 1989. *Humans and Other Animals*. London: Pluto Press.

O'Lemmon M. 2022. "The Worst Mistake 2.0? The Digital Revolution and the Consequences of Innovation," *AI & Society* [online first]. https://doi.org/10.1007/s00146-022-01599-5

Ord T. 2020. *The Precipice*. London: Bloomsbury.

Oswald A. & R. Winkelmann 2019. "Lottery Wins and Satisfaction: Overturning Brickman in Modern Longitudinal Data on Germany," in M. Rojas (Ed.), *The Economics of Happiness: How the Easterlin Paradox Transformed Our Understanding of Well-Being and Progress*. Cham: Springer.

Pareto V. 1916. *Treaty of General Sociology*. Florence: G. Barbéra.

Parfit D. 1984. *Reasons and Persons*. Oxford: Clarendon Press.

Parfit D. 2011. *On What Matters*. Oxford: Oxford University Press.

Paul II, John. 1994. *Letter to Families*. Vatican City: Libreria Editrice Vaticana.

Perkins M. A. 2004. *Christendom and European Identity*. Berlin: De Gruyter.

Rowe W. L. 1979. "The Problem of Evil and Some Varieties of Atheism," *American Philosophical Quarterly* 16(4):335–341.

Russell S. J. 2019. *Human Compatible*. New York: Viking.

Sahlins M. 1972. *Stone Age Economics*. New York: Routledge.

Shermer M. 2017. "Romance of the Vanished Past," *Scientific American* 316(6):75.

Singer P. 1972. "Famine, Affluence, and Morality," *Philosophy and Public Affairs* 1(3):229–243.

Singer P. 1975. *Animal Liberation*. New York: HarperCollins.

Singer P. 1981. *The Expanding Circle*. Oxford: Clarendon Press.

Singer P. 2009. *The Life You Can Save*. New York: Random House.

Singer P. 2015. *The Most Good You Can Do*. New Haven: Yale University Press.

Skutch A. 1948. "Life's Greatest Evil," *Scientific Monthly* 66(6):514–518.

Stevenson B. & J. Wolfers. 2008. "Economic Growth and Subjective Well-Being: Reassessing the Easterlin Paradox," *Brookings Papers on Economic Activity* 1:1–87.

Spinoza Benedictus de 1677. *Ethics*. Amsterdam: Jan Rieuwertsz.

Tennyson A. 1850. *In Memoriam*. London: Edward Moxon.

Tillich P. 1948. *The Protestant Era*. Chicago: University of Chicago Press.

Todd B., S. Farquhar, & P. Mills 2012. "The Ethical Careers Debate," *Oxford Left Review* 7:4–9.

Tu Weiming, M. Hejtmanek, & A. Wachman 1992. *The Confucian World Observed*. Honolulu: East-West Center.

Turing A. 1950. "Computing Machinery and Intelligence," *Mind* 59(236):433–460. https://doi.org/10.1093/mind/LIX.236.433

Unger P. 1993. *Living High and Letting Die*. Oxford: Oxford University Press.

Woods Jr., T. E. 2012. *How the Catholic Church Built Western Civilization*. Washington: Regnery.

Wright R. 2004. *A Short History of Progress*. Toronto: House of Anansi Press.

Zhang Shanruo Ning. 2016. *Confucianism in Contemporary Chinese Politics*. Lanham: Lexington Books.

Zuber S., N. Venkatesh, T. Tännsjö, Ch. Tarsney, H. Orri Stefánsson, K. Steele, D. Spears, J. Sebo, M. Pivato, T. Ord, Yew-Kwang Ng, M. Masny, W. MacAskill, N. Lawson, K. Kuruc, M. Hutchinson, J. E. Gustafsson, H. Greaves, L. Forsberg, M. Fleurbaey, D. Coffey, Susumu Cato, C. Castro, T. Campbell, M. Budolfson, J. Broome, A. Berger, N. Beckstead, & G. B. Asheim 2021. "What Should We Agree on about the Repugnant Conclusion?" *Utilitas* 33(4):379–383. https://doi.org/10.1017/S095382082100011X