The scope of longtermism

David Thorstad (Global Priorities Institute)

Global Priorities Institute | June 2021

GPI Working Paper No . 6-2021

The scope of longtermism

Abstract

Longtermism is the thesis that in a large class of decision situations, the best thing we can do is what is best for the long-term future. The *scope question* for longtermism asks: how large is the class of decision situations for which longtermism holds? In this paper, I suggest that the scope of longtermism may be narrower than many longtermists suppose. I identify a restricted version of longtermism: *swamping axiological strong longtermism* (swamping ASL). I identify three *scope-limiting factors* — probabilistic and decision-theoretic phenomena which, when present, tend to reduce the prospects for swamping ASL. I argue that these scope-limiting factors are often present in human decision problems, then use two case studies from recent discussions of longtermism to show how the scope-limiting factors lead to a restricted, if perhaps nonempty, scope for swamping ASL.

1 Introduction

If we play our cards right, the future of humanity will be vast and flourishing. The earth will be habitable for at least another billion years. During that time, we may travel well beyond the earth to settle distant planets. And increases in technology may allow us to live richer, longer and fuller lives than many of us enjoy today.

If we play our cards wrong, the future may be short or brutal. Already as a species we have acquired the capacity to make ourselves extinct, and many authors put forward alarmingly high estimates of our probability of doing so (Bostrom 2002; Leslie 1996; Ord 2020). Even if we survive long into the future, technological advances may be used to breed suffering and oppression on an unimaginable scale (Sotala and Gloor 2017; Torres 2018).

Some authors have taken these considerations to motivate *longtermism*: roughly, the thesis that in a large class of decision situations, the best thing we can do is what is best for the long-term future (Beckstead 2013; Greaves and MacAskill 2021; Greaves et al. forthcoming; MacAskill 2022; Ord 2020). The *scope question* for longtermism asks: how large is the class of decision situations for which longtermism holds?

Longtermism was originally developed to describe the decisions facing present-day philanthropists. Longtermists suggest that the best thing philanthropists can do today is to safeguard the long-term future. But many have held that the scope of longtermism extends considerably further. Hilary Greaves and Will MacAskill (2021) suggest that longtermism holds in all of the most important decisions facing humanity today. Nick Beckstead (2013) and Andreas Mogensen (2021) suggest that longtermism extends into global health decisionmaking. And Owen Cotton-Barratt (2021) suggests that even most mundane decisions, such as selecting topics for dinner-table conversation, should be made to promote proxy goals which track far-future value.

In this paper, I argue that the scope of longtermism may be narrower than many longtermists suppose. Section 2 clarifies my target: *ex ante*, swamping axiological strong longtermism (swamping ASL). Section 3 illustrates a historical decision problem in which swamping ASL may have been true. However, Sections 4-6 develop three scope-limiting factors: probabilistic and decision-theoretic phenomena which, when present, tend to reduce the prospects for swamping ASL. I argue that these scope-limiting factors are present in many human decision problems. Sections 7-8 use a pair of case studies to show how the presence of these scope-limiting factors leads to a limited, but perhaps nonempty, scope for swamping ASL. Section 9 concludes.

2 Preliminaries

2.1 Longtermism: axiological and ex ante

Longtermism comes in both axiological and deontic varieties. Roughly speaking, *axiological longtermism* says that the best options available to us are often near-best for the long-term future, and *deontic longtermism* says that we often should take some such option. Longtermists standardly begin by arguing for axiological longtermism, then arguing that axiological longtermism implies deontic longtermism across a wide range of deontic assumptions. In order to avoid complications associated with the passage between axiological and deontic claims, I focus on axiological rather than deontic longtermism.

Axiological longtermism can be construed as an *ex ante* claim about the values which options have from an *ex ante* perspective, or as an *ex post* claim about the value that options will in fact produce. It is generally thought that *ex post* longtermism is more plausible than *ex ante* longtermism, since many of our actions may in fact make a strong difference to the course of human history, even if we are not able to foresee what that difference will be.[1] For this reason, most scholarly attention has focused on *ex ante* versions of longtermism, and I follow this trend here.

The best-known view in this area is what has been called axiological strong longtermism (ASL):

> **(ASL)** In a wide class of decision situations, the option that is *ex ante* best is contained in a fairly small subset of options whose *ex ante* effects on the very long-run future are best.[2]

My target in this paper will be a restricted form of ASL.

2.2 Swamping axiological strong longtermism

Let a *longtermist option* be an option whose *ex ante* effects on the very long-run future are near-best.[3] ASL holds whenever the *ex ante* best option is a longtermist option. This can happen in two ways.

[1] However, Section 4 and on some views also Section 6 will place limits on the scope of *ex post* longtermism.

[2] This is the form of longtermism considered in Greaves and MacAskill (2019). Greaves and MacAskill (2021) defend a scope-restricted version of ASL, focusing only on the most important decision situations facing humanity today. I use the older, more general formulation of ASL in order to avoid ruling out wider scopes for ASL, and indeed Greaves and MacAskill are sympathetic to the idea that ASL has fairly wide scope.

[3] More formally, suppose that value is temporally separable, so that $V_o = S_o + L_o$ where V_o, S_o, L_o are the overall, short-term and long-term values of option o. Assess changes in value $\Delta V_o, \Delta S_o, \Delta L_o$ relative to a baseline, such as the effects of inaction. And take an expectational construal of *ex ante* value. Then a *longtermist option* is such that $E[\Delta L_o] \geq T * max_{o' \in O} E[\Delta L_{o'}]$ where O are the options available to the actor and T is a context-independent threshold for effects that count as 'near-best'. Perhaps we might take $T = 0.9$.

First, let a *swamping option* be an option whose expected long-term benefits exceed in magnitude the expected short-term effects produced by any option.[4] I call these swamping options because their long-term effects begin to swamp short-term considerations in determining *ex ante* value. The first way for ASL to be true is if the best option is both a longtermist option and a swamping option.

> **Swamping axiological strong longtermism (Swamping ASL)** In a wide class of decision situations, the option that is *ex ante* best is a swamping longtermist option.

My focus in this paper will be on swamping ASL.

Second, the best option may be a *non-swamping longtermist option*, an option whose expected long-term effects are near-best, but do not exceed in magnitude the expected short-term effects of all other options. One way to defend the value of non-swamping longtermist options would be through the *convergence thesis* that what is best for the short-term is often near-best for the long-term as well.[5] The convergence thesis suggests that even when long-term effects do not swamp short-term effects in magnitude, the best option may nonetheless be a longtermist option, since the best short-term option will often be near-best for the long-term.

I focus on swamping ASL for three reasons. First, swamping ASL figures in leading philosophical arguments for ASL and in most nonphilosophical treatments of longtermism. Second, swamping ASL is the most distinct and revisionary form of ASL, because it tells us that the short-termist options we might have assumed to be best are in fact often not best.[6] Third, swamping ASL underlies many of the most persuasive arguments

[4] Using the notation and assumptions of the previous footnote, a *swamping longtermist* option is such that $E[\Delta L_o] > max_{o' \in O} |E[\Delta S_{o'}]|$ where O are the options available to the actor. This is a simplification of the model from Greaves and MacAskill (2019).

[5] For example, you might think that the best thing we can do to ensure a good future is to promote economic growth (Cowen 2018), and that is also among the best things we can do for the short-term. Note, however, that this may be an example of a swamping longtermist option.

[6] Strictly speaking, this does not follow from swamping ASL since swamping ASL is compatible with the convergence thesis. However, in practice most of the examples used to support swamping ASL are not near-best in their short-term effects.

from axiological to deontic longtermism, which rely on the claim that sufficiently strong duties to promote impartial value may trump competing nonconsequentialist duties. As we move away from swamping longtermism, obligations to promote long-term value will diminish in strength, putting pressure against the inference from axiological to deontic longtermism.

2.3 Scope-limiting phenomena

In this paper, I illustrate three *scope-limiting phenomena*. These are probabilistic and decision-theoretic phenomena which, when present in a decision problem, tend to reduce the prospects for swamping ASL to hold in that problem. Sections 4-6 introduce the scope-limiting phenomena that will concern me: rapid diminution (Section 4); washing out (Section 5); and option unawareness (Section 6). I argue that each scope-limiting phenomenon is often present in the decisions that we face, then show how the presence of each phenomenon reduces the prospects for swamping ASL.

To say that these scope-limiting phenomena reduce the prospects for swamping ASL is not to say that the swamping ASL has empty scope. Section 3 illustrates a case in which swamping ASL may well have been true, and Section 7 argues that this case is not significantly afflicted by any of the scope-limiting phenomena. Moreover, it is not impossible for swamping ASL to hold in some cases where all of the scope-limiting phenomena obtain. However, the presence of these scope-limiting phenomena does put pressure on many cases in which swamping ASL has been claimed to obtain. Section 8 illustrates one case of this type.

Summing up, my target in this paper is *ex ante*, swamping axiological strong longtermism. I illustrate three scope-limiting phenomena to suggest that swamping ASL has more limited scope than we might otherwise suppose. But first, let us consider where swamping ASL may be plausible.

3 Swamping ASL and the Space Guard Survey

A popular way to motivate swamping ASL is to think about risks of human extinction (Bostrom 2013; Greaves and MacAskill 2021; Ord 2020). Now on some views, the continued survival of humanity may have indifferent, or even negative value (Benatar 2006). Given our potential to spread death and suffering, the universe may be better off once it is rid of humanity. On these views, risks of human extinction will not motivate swamping ASL. But many philosophers are cautiously optimistic that the survival of humanity would be a good thing (Beckstead 2013; Ord 2020; Parfit 2011). On these views, it may be very important to protect humanity from premature extinction. And in some cases, decisions to mitigate extinction risk may motivate swamping ASL.

One way that humans might go extinct is through the impact of a large asteroid on earth. Indeed, there is mounting evidence that an asteroid impact during the Cretaceous period killed every land-dwelling vertebrate with mass over five kilograms (Alvarez et al. 1980; Schulte et al. 2010). As recently as 2019, an asteroid 100 meters in diameter passed five times closer to the earth than the average orbital distance of the moon and was detected only a day before it arrived (Zambrano-Marin et al. 2021).

NASA classifies asteroids with diameter greater than 1 kilometer as catastrophic, capable of causing a global calamity or even mass-extinction. Our best estimates suggest that such impacts occur on earth about once in every 6,000 centuries (Stokes et al. 2017). Plausibly, it is worth our while to detect and prepare for such events.

As evidence mounted of the threat posed by asteroid impacts, the United States Congress funded the Space Guard Survey, a collection of projects aimed at tracking potentially dangerous asteroids, comets and other near-earth objects. Since the 1990s, the Space Guard Survey has mapped approximately 95% of the near-earth asteroids with diameters exceeding 1 kilometer, at a cost of $70 million. From an *ex ante* perspective, how valuable was the Space Guard Survey?

Let us work with a set of conservative assumptions, so we cannot be accused of rigging the numbers. Assume first that the Space Guard Survey can only accurately predict impacts during the next century. Next, suppose that if an undetected asteroid with diameter greater than 1 kilometer were to strike earth during the next century, the chance of extinction would be one in a million. Now, consider that estimates of the expected number of future humanlike lives range from about 10^{13} to 10^{55} (Bostrom 2014; Newberry 2021). This puts the Space Guard Survey's expected cost of detecting an extinction-causing asteroid impact, counting only impacts within the next century, at about $7 per expected future life, and fractions of a penny using anything but the most conservative estimate of future lives.[7] For comparison, our best estimates put the cost of saving a life through short-termist interventions at several thousand dollars (GiveWell 2021), far exceeding the cost of the Space Guard Survey if we have any confidence at all in our ability to prepare for and survive an otherwise-catastrophic impact with sufficient warning.

Now consider the decision facing Congress in the early 1990s: whether to fund the Space Guard Survey or to redirect the money towards alternative programs. Suppose, plausibly, that the expected long-term effects of the Space Guard Survey were near-best out of all programs available for Congress to fund. Or, if this is not plausible, replace the Space Guard Survey with any program that had near-best expected long-term effects and repeat the argument. Then suppose we also grant that the expected long-term effects of the Space Guard Survey exceeded in magnitude the best-achievable short-term effects of any competing program. For example, we might benchmark the long-term effects of the Space Guard Survey at several dollars per life saved, and the best-achievable short-term effects of competing programs at several thousand dollars per life saved. If this is right, then swamping ASL was true of Congress's decision problem. Funding the Space Guard Survey was the best thing that Congress could have done; its long-term effects were near-best, and they swamped in magnitude the expected short-term effects of all

[7] This estimate is arrived at by multiplying the expected number of future lives by the per-century probability of a catastrophic asteroid impacting earth, as well as by the probability that an undetected catastrophic asteroid impact would lead to extinction, then dividing the result by the program cost.

options. Indeed, it may be precisely on these grounds that Congress decided to fund the Space Guard Survey.

Some readers might disagree with the claim that swamping ASL holds of Congress's decision problem. Perhaps you hold a person-affecting axiology on which it is neither good nor bad to ensure that future humans come into existence. Or perhaps you think that the likely outcome of asteroid detection research is research into dangerous technologies for asteroid deflection, and that the dangers posed by these technologies are greater than the dangers they eliminate (Ord 2020). But in this paper, I want to emphasize a different line of resistance: cases such as the Space Guard Survey are quite special (Section 7), in that they avoid a number of scope-limiting phenomena (Sections 4-6) that serve to reduce the prospects for swamping ASL. This means that we can, and perhaps should, acknowledge some cases in which swamping ASL holds, while resisting swamping ASL as a description of many other decision problems.

4 Rapid diminution

In the next three sections, I illustrate a series of scope-limiting factors. I argue that these factors are often present in the decisions that we face and that, when present, these factors tend to reduce the prospects for swamping ASL.

The first scope-limiting factor is *rapid diminution*. Fix an option o and consider the probability distribution over long-term impacts of o.[8] In most cases, the probabilities of long-term impacts decrease as those impacts increase in magnitude. If probabilities of impacts decrease more slowly than the magnitudes of those impacts increase, then the expected long-term consequences of o may be astronomically high. But if the probabilities of large impacts decrease quickly, the expected long-term impacts of o may be quite modest.

Rapid diminution is a familiar feature of many of the best-known probability distri-

[8] I.e. consider the probability distribution over the partition $\{[\Delta L = k] : k \in \mathbb{R}\}$.

butions. For example, suppose that we model the expected long-term impact of *o* using a normal distribution, centered around the origin, with a standard deviation equivalent to the value of ten lives saved. On this model, the probability of long-term impacts exceeding five times this value is less than one in a million. And the probabilities of astronomical long-term impacts, while nonzero, will be so negligible as to have no significant impact on the expected long-term impact of *o*.

The argument from rapid diminution claims that many options exhibit rapid diminution in the probability of long-term impacts, limiting the contribution that long-term impacts can make to the expected value of those options. This argument is supported by *persistence skepticism*: the view that many of our actions do not make a large persisting impact on the long-term future.

We can assess the case for persistence skepticism by looking at the burgeoning academic field of persistence studies, which studies examples of persistent long-term changes (Alesina and Giuliano 2015; Nunn 2020). Persistence studies often returns surprising negative results, where effects that we might have expected to persist for a long time evaporate after several decades. For example, given the scale of American bombing in Japan and Vietnam, one might expect persistent economic effects in the heaviest-hit areas. Given the number of people affected and the magnitude of potential effects, this is exactly the type of persistent effect that would interest a longtermist. But a half-century later, there are no statistically significant differences between the most- and least-affected areas on standard economic indicators such as population size, poverty rates and consumption patterns (Davis and Weinstein 2008; Miguel and Roland 2011). For a striking example, the cities of Hiroshima and Nagasaki returned to their pre-war population levels by the mid-1950s.

Now it is true that persistence studies has identified a few-dozen effects which might be more persistent. For example, the introduction of the plough may have affected fertility norms and increased the gendered division of labor (Alesina et al. 2011, 2013); the African slave trade may have stably reduced social trust and economic indicators in the hardesthit regions (Nunn 2008; Nunn and Wantchekon 2011); and the Catholic Church may be responsible for the spread of so-called WEIRD personality traits identified by comparative psychologists (Schulz et al. 2019). However, these findings need to be taken with three grains of salt.

First, many of these findings are controversial, and alternative explanations have been proposed (Kelly 2019; Sevilla 2021). Second, these findings are few and far between, so together with other negative findings they may not challenge the underlying rarity of strong long-term effects. And finally, most of the examples in this literature also involve short-term effects of comparable importance to their claimed long-term effects. Hence the persistence literature may not provide strong support for the swamping longtermist's hope that persistent long-term effects could swamp short-term effects in importance.

At the same time, there is no doubt that some actions have a nontrivial probability of making persistent changes to the value of the future far greater than any of their short-term effects. As a result, we cannot get by with the argument from rapid diminution alone. We need to supplement rapid diminution with a second scope-limiting factor: washing out.

5 Washing out

A second scope-limiting factor is *washing out*. Although many options have nontrivial probabilities of making positive impacts on the future, they also have nontrivial probabilities of making negative impacts. For example, by driving down the road I might crash into the otherwise-founder of a world government, but I might also crash into her chief opponent. As a result, the argument from washing out holds that there will often be significant cancellation between possible positive and negative effects in determining the expected values of options.

There are two related ways that the argument from washing out can be articulated. The first begins with the popular Bayesian idea that complete ignorance about the long-term value of an option should be represented by a symmetric prior distribution over possible long-term values. Next, the argument notes that we are often in a situation of *evidential paucity*: although we have some new evidence bearing on long-term values, often our evidence is quite weak and undiagnostic. As a result, the prior distribution will exert a significant influence on the shape of our current credences, so if the prior is symmetric then our current credences should be fairly symmetric as well. And a near-symmetric probability distribution over long-term impacts gives significant cancellation when we take expected values.

We can make a similar point by arguing for *forecasting pessimism*, the view that it is often very difficult to predict the impact of our actions on far-future value. For example, there is no doubt that the Roman sacking of Carthage had a major impact on our lives today, by cementing the Roman empire and changing the course of Western civilization. But even today, let alone with evidence available at the time, it is very difficult to say whether that impact was for good or for ill.

Forecasting pessimism generates a type of washing out between possible positive and negative forecasts.[9] When we make forecasts based on sparse data, we need to take account of the fact that the data we have been dealt is a noisy reflection of the underlying reality. As phenomena become more unpredictable and our data becomes increasingly sparse, we should grow more willing to chalk up any apparent directionality in our forecasts to noisiness in the hand of data that nature has dealt us. In other words, as forecasting becomes more difficult we get increasing wash-out between possible positive and negative forecasts that we could have made had nature dealt us different samples of data.

Why should we be pessimistic about our ability to forecast long-run value? Intuitions about the sacking of Carthage are well and good, but it would be nice to have some concrete theoretical considerations on the table. Here are three reasons to think that we are often in a poor position to forecast long-run value.

[9] Among the many ways to give formal expression to this idea, Gabaix and Laibson's (2021) as-if discounting brings out the similarity to the argument from evidential paucity by highlighting the role of priors.

First, we have limited and mixed *track records* of making long-term value forecasts. We do not often make forecasts even on a modest timeline of 20-30 years, and as a result there are only a few studies assessing our track record at this timescale.[10] These studies give a mixed picture of our track record at predicting the moderately-far future: in some areas our predictions are reasonably accurate, whereas in others they are not. But the longtermist is interested in predictions at a timescale of centuries or millennia. We have made and tested so few predictions at these time scales that I am aware of no studies which assess our track record at this timescale outside of highly circumscribed scientific domains, and if our moderate-future track record is any indication, our accuracy may decline quite rapidly this far into the future.

Second, there is an enormous amount of *practitioner skepticism* on behalf of prominent academic and non-academic forecasters about the possibility of making forecasts on a timescale of centuries, particularly when we are interested in forecasting rare events, as longtermists often are. Very few economists, risk analysts, and other experts are willing to make such predictions, citing the unavailability of data, a lack of relevant theoretical models, and the inherent unpredictability of underlying systems (Freedman 1981; Goodwin and Wright 2010; Makridakis and Taleb 2009). And when risk analysts are asked to consult on the management of very long-term risks, they increasingly apply a variety of non-forecasting methods which enumerate and manage possible risks without any attempt to forecast their likelihood (Marchau et al. 2019; Ranger et al. 2013). If leading practitioners are unwilling to make forecasts on this timescale and increasingly suggest that we should act without forecasting, this is some evidence that the underlying phenomena may be too unforeseeable to effectively forecast.

Third, *value is multidimensional*. The value of a time-slice in human history is determined by many factors such as the number of people living, their health, longevity, education, and social inclusion. It is often relatively tractable to predict a single quantity,

[10] For domain-specific track records see Albright (2002); Kott and Perconti (2018); Parente and Anderson-Parente (2011); Risi et al. (2019) and Yusuf (2009). For discussion see Fye et al. (2013) and Mullins (2018).

such as the number of malaria deaths that will be directly prevented by a program of distributing bed nets. And when we assess the track records of past predictions, we often assess predictions of this form. But the longtermist is interested in predicting value itself, which turns on many different quantities. This is harder to predict: distributing bed nets also affects factors such as population size, economic growth, and government provision of social services (Deaton 2015). So even if we think that the long-term effects of a program along a single dimension of value are fairly predictable, we may think that the ultimate value of the intervention is much less predictable.

Summing up, the argument from washing out claims that we often get significant cancellation between possible positive and negative effects of an intervention when taking expected values. One window into washing out comes from evidential paucity: because we have little evidence about long-term impacts, we should adopt a fairly-symmetric probability distribution over possible long-term impacts. The same phenomenon occurs in thinking about forecasting. Because our evidence about far-future value is sparse, we should think that our forecasts could easily have been different if we had received different evidence about the future, and as a result we get significant cancellation between possible positive and negative forecasts of far-future value.

Together, rapid diminution and washing out put pressure on the scope of swamping ASL. They do this by suggesting that the expected far-future benefits of many options may be relatively modest, and may be significantly cancelled by the expected far-future costs of these options. In the next section, I illustrate a third and final scope-limiting factor: option unawareness.

6 Option unawareness

Rational *ex ante* choice involves taking the *ex ante* best option from the options available to you. But which options are these? We might take a highly unconstrained reading on which any option that is physically possible to perform belongs to your choice set. But in practice, this reading seems to betray the *ex ante* perspective (Hedden 2012).

Suppose you are being chased down an alleyway by masked assailants. A dead end approaches. Should you turn right, turn left, or stop and fight? Trick question! I forgot to mention that you see a weak ventilation pipe which, if opened, would spray your attackers with hot steam. That's better than running or fighting. Let us suppose that, in theory, all of this could be inferred with high probability from your knowledge of physics together with your present perceptual evidence, but you haven't considered it. Does this mean that you would act wrongly by doing anything except breaking the pipe?

Many decision theorists have thought you would not act wrongly here. Just as *ex ante* choosers have limited information about the values of options, so too they have limited awareness of the many different options in principle available to them. Theories of *option unawareness* incorporate this element of *ex ante* choice by restricting choice sets to options which an agent is, in some sense, relevantly aware of (Bradley 2017; Karni and Vierø 2013; Steele and Stefánsson 2021). In the present case, this means that your options are first described: turning right, turning left, or stopping to fight. Unless, perhaps, you happen to be James Bond.

How is option awareness relevant to swamping ASL? To see the relevance, note that rapid diminution and washing out are features of options, not decision problems. Together, rapid diminution and washing out imply that many of the options we face will not be swamping longtermist options, because their expected far-future benefits may be relatively modest and may be significantly cancelled by expected far-future costs. However, swamping ASL is a thesis about decision problems, which present us with a set of options rather than a single option. Swamping ASL holds in any decision problem for which the *ex ante* best option is a swamping longtermist option. The presence of a single swamping longtermist option in a decision problem may be enough to vindicate swamping ASL.

This means that the number of options present in a decision problem bears strongly on the likelihood that swamping ASL will be true in that problem. If the vast majority of options are not swamping longtermist options, then swamping ASL will be unlikely to hold in decision problems containing a dozen options, since it is unlikely that any of these will be swamping longtermist options. But swamping ASL may be more likely to hold in decision problems containing millions or billions of options, simply because one of those options is likely to be a swamping longtermist option, and because swamping longtermist options are often, when present, the best options we can take.[11] Hence swamping ASL may be relatively plausible before we restrict agents' option sets to incorporate their limited awareness of available options, but less plausible once option unawareness is incorporated.

To see the point in context, consider interventions aimed at combatting childhood blindness. Nick Beckstead (2013) has suggested that the short-term benefits of these interventions, namely preventing children from going blind, may be swamped by the long-term benefits of preventing blindness, such as speeding up a nation's economic development or changing the world's trajectory by changing the role that children will play in the national and global economy. Our discussion of rapid diminution and washing out suggests that, for most particular children, Beckstead's claim will be false. Because it is hard for a single individual to make a lasting impact on the long-term future, and because individuals may also make negative impacts on the long-term future, for most children, the expected benefit of preventing them from going blind will be driven primarily by short-term considerations, such as the value of not being blind.

However, perhaps it is not implausible that somewhere in the world, there is a collection of seventeen children and a sequence of days such that, if each child were given preventative treatment on the requisite day, the long-term trajectory of the world would be significantly improved. Let O^* be the option of giving just this course of treatment to each of the children in question. And perhaps it is not unreasonable to suppose that,

[11] As always, there is a problem of option individuation, since it is often possible to chop a single option into millions or billions of nearly-identical options, but that is unlikely to improve the prospects of swamping ASL. Readers are invited to approach this discussion in a way that treats awareness of *relevantly different* options as raising the prospects for swamping ASL to be true. Like most philosophers, I do not pretend to be in possession of a formal criterion for relevant difference, or another fully formal solution to the problems induced by option individuation.

in principle, the high value of O^* could be worked out *ex ante* on the basis of available information, even if the calculations required to see this would be astronomically complex.

Now suppose that you have five thousand dollars to spend, and you want to use that money to combat childhood blindness. We might take an awareness-restricted view of your decision problem, on which you are deciding among donating to the half-dozen most prominent international efforts to combat childhood blindness. In this problem, swamping ASL may be relatively implausible. On the other hand, we might take an awareness-unrestricted view of your decision problem, on which you are deciding among any physically possible use of five thousand dollars to combat childhood blindness, including options such as O^*. In this awareness-unrestricted decision problem, swamping ASL may be more plausible. In this way, the prospects for swamping ASL may be substantially reduced once reasonable levels of option unawareness are incorporated into *ex ante* decisionmaking.

So far, we have met three scope-limiting factors: rapid diminution, washing out, and option unawareness. We saw that these scope-limiting factors are often present in decisionmaking, and that, when present, they tend to diminish the prospects for swamping ASL. But this does not imply that swamping ASL has empty scope. To see the point, let us return to our discussion of the Space Guard Survey.

7 The good case revisited

In Section 3, I argued that swamping ASL may have accurately described a decision problem facing Congress in the 1990s: whether to fund the Space Guard Survey, or to redirect the money elsewhere. In support of that suggestion, note that all three of the scope-limiting factors introduced above are largely absent from this example.

Begin with the problem of rapid diminution: the probabilities of large long-term impacts diminish rapidly. The argument for rapid diminution drew on skepticism about the persistence of short-term effects into the long-term future. It is often hard to make a persisting impact on the long-term future. But it is not hard to see how the proposed effects of asteroid detection, namely preventing human extinction, could persist into the long-term future.[12] Not being extinct is a status that can last for a very long time if we play our cards right.

Turn next to the problem of washing out: possible long-term benefits may be significantly cancelled by possible long-term harms. The first argument for washing out drew on evidential paucity: we don't have much evidence about the long-term effects of our actions. But asteroid detection is an area in which we do have significant evidence about possible long-term effects. This includes evidence from past asteroid impacts together with a good scientific understanding of the determinants of asteroid impact force, which is sufficient to build compelling computational models of impact damages (Stokes et al. 2017).

Our second argument for washing out drew on forecasting skepticism: it is hard to predict the future. First, I argued that in many areas we have no good track record of predicting the far future. But astronomy is one of the few areas in which we have a good track record of predictions on this time-scale. Second, I argued that experts are often unwilling to make forecasts of the relevant type. But the key forecast driving the example was a prediction by NASA scientists of the probability of catastrophic asteroid impacts. Third, I argued that due to the multidimensionality of value we may only be able to estimate the probability of a catastrophic impact, but not its value. But where human extinction is concerned, this may not be a significant problem. To evaluate whether preventing human extinction would be a good thing, we must only answer a single question: whether the continued existence of humanity would be a good thing. While answering this question is not straightforward, many theorists are cautiously optimistic that the future will be good (Beckstead 2013; Ord 2020; Parfit 2011).

Turn finally to the problem of option unawareness: decisionmakers are unaware of

[12] However, if we are pessimistic about current levels of existential risk, this point is no longer so clear (Thorstad 2022).

some options which may be swamping longtermist options. But in the case of the Space Guard Survey, we were already aware of feasible options which could produce the desired results at a reasonable cost. It may well be true that other options, of which we were unaware, would have been still better, but this does not mean that the options ultimately chosen were not swamping longtermist options.

So far, we have seen that the scope-limiting factors do not threaten the case for swamping ASL in some cases, for example the decision to fund the Space Guard Survey. That should be unsurprising: we did not expect the scope of swamping ASL to be completely empty, and the Space Guard Survey is an example in which many decisionmakers agreed with the longtermist's evaluative claims. However, in many other examples the scope-limiting factors begin to significantly threaten the case for swamping ASL. The next section provides an illustration.

8 Beyond the good case

Let us return to Beckstead's case of a philanthropist deciding between various initiatives for preventing childhood blindness. We have already seen that this case is subject to significant option unawareness, and that the presence of option unawareness tends to reduce the plausibility of swamping ASL in this case. In the rest of this section, I suggest that both of the remaining scope-limiting factors are also present in this case, and that these factors further tell against the applicability of swamping ASL.

Begin with rapid diminution in the probabilities of large long-term impacts. The argument for rapid diminution was that it is hard to make a persisting impact on the long-term future. For example, Beckstead suggests that curing blindness may impact the long-term future by helping treated individuals to contribute to their nation's economic development. But we saw in Section 4 that even large shocks, such as the detonation of a thermonuclear bomb, are often insufficient to make lasting long-term impacts on the economy of a medium-sized city, much less a nation. If that is right, then we should substantially reduce our confidence in the ability of any single individual to make a persisting long-term economic impact. It is true, of course, that some individuals may occupy prominent economic roles, for example the leadership of a large corporation. But what is less clear is that, in the absence of these individuals, underlying demographic, cultural and economic factors would have led the region down a substantially different path.

Turn next to washing out: the tendency for long-term expected benefits to be significantly cancelled by long-term expected harms. I think that we should expect significant washing out in this case. It is, of course, quite possible that the children we treat will go on to fight climate change or found a world government. But it is also possible that they will go on to be among the world's greatest polluters, or to oppose world government. Nor, for that matter, can we be terribly certain which of these developments would be for the long-term good. It might be that the premature move towards world government would lead to tyranny, or to a governance failure that would set back the development of more effective systems by several centuries. And for that matter, we should not be terribly confident that blindness will be an impediment to playing an important role in any of these endeavors. Because we have very little evidence to go on in assessing the likelihood of various far-future effects that may result from treating childhood blindness, we should tend to significantly discount likely long-term benefits by leaving open the real possibility that our actions will produce long-term harms.

The discussion of childhood blindness helps us to see how quickly the scope-limiting factors get a take on decisionmaking, even in cases that are often taken to motivate swamping ASL. When the scope-limiting factors are present, the case for swamping ASL becomes much more tenuous.

9 Conclusion

This paper assessed the fate of *ex ante* swamping ASL: the claim that the *ex ante* best thing we can do is often a swamping longtermist option. I argued that swamping ASL may hold in some cases, such as the decision to fund the Space Guard Survey. However, I also discussed three *scope-limiting factors* which, when present in a decision problem, tend to reduce the prospects for swamping ASL. These scope-limiting factors included *rapid diminution* in the probabilities of large far-future benefits; *washing out* between possible positive and negative future effects; and *unawareness* of swamping longtermist options.

I argued that swamping ASL may still be true in some cases, particularly when the scope-limiting factors are not present. However, I suggested that the scope of swamping ASL may be far narrower than often supposed. I used a discussion of treating childhood blindness to illustrate how the scope-limiting factors get a take even on many cases taken to motivate swamping ASL. I suggested that as the scope-limiting factors make themselves increasingly felt, the prospects for swamping ASL diminish.

In some ways, this may be familiar and comforting news. For example, Hilary Greaves (2016) considers the cluelessness problem that we are often significantly clueless about the *ex ante* values of our actions because we are clueless about their long-term effects. Greaves suggests that although cluelessness may correctly describe some complex decisionmaking problems, we should not exaggerate the extent of *mundane cluelessness* in everyday decisionmaking. A natural way of explaining this result would be to argue that in most everyday decisionmaking, it is the expected long-term effects of our actions that are swamped by their short-term effects, and not the other way around. This would mean that cluelessness about long-term effects is often compatible with substantial confidence and precision in our views about the overall values of options.

In addition, this discussion leaves room for swamping ASL to be true and important in some contemporary decision problems. It also does not directly pronounce on the fate of ex-post versions of ASL, or on the fate of non-swamping ASL. However, it does suggest that swamping versions of ASL may have a more limited scope than otherwise supposed.

References

Albright, Richard. 2002. "What can past technology forecasts tell us about the future?" *Technological Forecasting and Social Change* 69:443–464.

Alesina, Alberto and Giuliano, Paola. 2015. "Culture and institutions." *Journal of Economic Literature* 53:898–944.

Alesina, Alberto, Giuliano, Paola, and Nunn, Nathan. 2011. "Fertility and the plough." *American Economic Review* 101:499–503.

—. 2013. "On the origins of gender roles: Women and the plough." *Quarterly Journal of Economics* 128:469–530.

Alvarez, Luis W., Alvarez, Walter, Asaro, Frank, and Michel, Helen V. 1980. "Extraterrestrial cause for the Cretaceous-Tertiary extinction." *Science* 208:1095–1180.

Beckstead, Nicholas. 2013. *On the overwhelming importance of shaping the far future*. Ph.D. thesis, Rutgers University.

Benatar, David. 2006. *Better never to have been: The harm of coming into existence*. Oxford University Press.

Bostrom, Nick. 2002. "Existential risks: Analyzing human extinction scenarios and related hazards." *Journal of Evolution and Technology* 9:1–30.

—. 2013. "Existential risk prevention as a global priority." *Global Policy* 4:15–31.

—. 2014. *Superintelligence*. Oxford University Press.

Bradley, Richard. 2017. *Decision theory with a human face*. Cambridge University Press.

Cotton-Barratt, Owen. 2021. "Everyday longtermism." EA Forum. https://forum.effectivealtruism.org/posts/3PmgXxBGBFMbfg4wJ/everyday-longtermism.

Cowen, Tyler. 2018. *Stubborn attachments*. Stripe Press.

Davis, Donald and Weinstein, David. 2008. "A search for multiple equilibria in urban industrial structure." *Journal of Regional Science* 48:29–62.

Deaton, Angus. 2015. *The great escape: Health, wealth, and the origins of inequality*. Princeton University Press.

Freedman, David. 1981. "Some pitfalls in large econometric models: A case study." *Journal of Business* 54:479–500.

Fye, Shannon, Charbonneau, Steven, Hay, Jason, and Mullins, Carie. 2013. "An examination of factors affecting accuracy in technology forecasts." *Technological Forecasting and Social Change* 80:1222–1231.

Gabaix, Xavier and Laibson, David. 2021. "Myopia and discounting." National Bureau of Economic Research Working Paper 23254.

GiveWell. 2021. "GiveWell's Cost-Effectiveness Analyses." https://www.givewell.org/how-we-work/our-criteria/cost-effectiveness/cost-effectiveness-models.

Goodwin, Paul and Wright, George. 2010. "The limits of forecasting methods in anticipating rare events." *Technological Forecasting and Social Change* 77:355–368.

Greaves, Hilary. 2016. "Cluelessness." *Proceedings of the Aristotelian Society* 116:311–339.

Greaves, Hilary and MacAskill, William. 2019. "The case for strong longtermism." Global Priorities Institute Working Paper 7-2019.

—. 2021. "The case for strong longtermism." Global Priorities Institute Working Paper 5-2021.

Greaves, Hilary, Thorstad, David, and Barrett, Jacob (eds.). forthcoming. *Longtermism*. Oxford University Press.

Hedden, Brian. 2012. "Options and the subjective ought." *Philosophical Studies* 343–360.

Karni, Edi and Vierø, Marie-Louise. 2013. "'Reverse Bayesianism': A choice-based theory of growing awareness." *American Economic Review* 103:2790–2810.

Kelly, Morgan. 2019. "The standard errors of persistence." CEPR Discussion Papers 13783.

Kott, Alexander and Perconti, Phillip. 2018. "Long-term forecasts of military technologies for a 20-30 year horizon: An empirical assessment of accuracy." *Technological Forecasting and Social Change* 137:272–9.

Leslie, John. 1996. *The end of the world: The science and ethics of human extinction*. Routledge.

MacAskill, William. 2022. *What we owe the future*. Basic books.

Makridakis, Spyros and Taleb, Nassim. 2009. "Decision making and planning under low levels of predictability." *International Journal of Forecasting* 25:716–733.

Marchau, Vincent, Walker, Warren, Bloemen, Pieter, and Popper, Steven (eds.). 2019. *Decision making under deep uncertainty*. Springer.

Miguel, Edward and Roland, Gérard. 2011. "The long-run impact of bombing Vietnam." *Journal of Development Economics* 96:1–15.

Mogensen, Andreas. 2021. "Maximal cluelessness." *Philosophical Quarterly* 71:141–62.

Mullins, Carie. 2018. "Retrospective analysis of long-term forecasts." Technical report, Open Philanthropy Project.

Newberry, Toby. 2021. "How cost-effective are efforts to detect near-Earth-objects?" Technical report, Global Priorities Institute.

Nunn, Nathan. 2008. “The long term effects of Africa’s slave trades.” *Quarterly Journal of Economics* 123:139–176.

—. 2020. “The historical roots of economic development.” *Science* 367:eaaz9986.

Nunn, Nathan and Wantchekon, Leonard. 2011. “The slave trade and the origins of mistrust in Africa.” *American Economic Review* 3221–3252.

Ord, Toby. 2020. *The precipice*. Bloomsbury.

Parente, Rick and Anderson-Parente, Janet. 2011. “A case study of long-term Delphi accuracy.” *Technological Forecasting and Social Change* 78:1705–1711.

Parfit, Derek. 2011. *On what matters*, volume 1. Oxford University Press.

Ranger, Nicola, Reeder, Tim, and Lowe, Jason. 2013. “Addressing ‘deep’ uncertainty over long-term climate in major infrastructure projects: Four innovations of the Thames Estuary 2100 project.” *EURO Journal on Decision Processes* 1:233–262.

Risi, Joseph, Sharma, Amit, Shah, Rohan, Connelly, Matthew, and Watts, Duncan. 2019. “Predicting history.” *Nature Human Behavior* 3:906–912.

Schulte, Peter et al. 2010. “The Chicxulub asteroid impact and mass extinction at the Cretaceous-Paleogene boundary.” *Science* 327:1214–1218.

Schulz, Jonathan F., Bahrami-Rad, Duman, Beauchamp, Jonathan, and Henrich, Joseph. 2019. “The Church, intensive kinship, and global psychological variation.” *Science* 36:eaau5141.

Sevilla, Jaime. 2021. “Persistence: A critical review.” Technical report, Forethought Foundation.

Sotala, Kaj and Gloor, Lukas. 2017. “Superintelligence as a cause or cure for risks of astronomic suffering.” *Informatica* 41:389–400.

Steele, Katie and Stefánsson, Orri. 2021. *Beyond uncertainty*. Cambridge University Press.

Stokes, Grant, Barbee, Brent, Bottke, William, et al. 2017. "Update to determine the feasibility of enhancing the search and characterization of NEOs: Report of the near-earth object science definition team." Technical report, NASA.

Thorstad, David. 2022. "Existential risk pessimism and the time of perils." Global Priorities Institute Working Paper 1-2022.

Torres, Phil. 2018. "Space colonization and suffering risks: Reassessing the 'maxipok rule'." *Futures* 100:74–85.

Yusuf, Moeed. 2009. "Predicting proliferation: the history of the future of nuclear weapons." Technical report, Brookings Institution.

Zambrano-Marin, L.F., Howell, E.S., Devogéle, M., et al. 2021. "Radar observations of near-earth asteroid 2019 OK." In *Proceedings of the 52nd Lunar and Planetary Science Conference 2021*, LPI Contribution Number 2548.