The scope of longtermism

1 The scope of longtermism

I A recent statement of longtermism:

> ***Ex ante axiological strong longtermism* (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. (Greaves and MacAskill ms).

Note: This is *ex ante* and axiological.

II **Scope question for ASL:** How wide is the class of decision situations satisfying ASL?

III **Current views:**

i **Opponents of longtermism:** It's an empty class.

ii **Cotton-Barratt:** Maximally wide:

> Strong longtermism says that the morally right thing to do is to make **all** decisions according to long-term effects.

Examples: Electing political candidates; teaching primary school children; managing workers; writing science journalism; choosing dinner-table conversation. (Cotton-Barratt 2021).

iii **Greaves and MacAskill:** Perhaps quite wide:

> **(Deworming)** Sophie is a philanthropic grantmaker comparing two deworming programs. ... The *short-term* cost-effectiveness of program A is higher than that of program B. ... However ... there are also effects on later life incomes, and thereby (presumably) on later-generation incomes. ... It could easily be the case that attending to the longer-term benefits of deworming reverses the comparative cost-effectiveness estimate of A vs. B. (Greaves and MacAskill ms).

iv **Beckstead:** Flirts with some similar cases:

> **(Blindness)** Suppose I cure some child's blindness. We ordinarily think that the main benefit of this is that the child will have a better life. ... However, curing the child's blindness creates a ripple effect that carries forward through many people affected by the child, and many people those people affect, and so forth. If those ripple effects continue for a long time, it can be argued that they swamp the proximate benefits that command our attention. (Beckstead 2013, p. 3).

IV **My target view:** The scope of longtermism is:

i Nonempty.

ii Contains the case of present-day, cause-neutral philanthropy and adjacent areas.

iii May thin out considerably beyond that.

V **Structure of argument:**

i Clarify the type of ASL that interests me.

ii Argue that ASL is plausible true in the good case of present-day, cause-neutral philanthropy.

iii Give two arguments for the **rarity thesis** that (roughly) the options needed to witness ASL in a given problem are rare.

iv Show that the rarity thesis does not threaten ASL in the good case.

v Use the rarity thesis to propose two challenges which limit the scope of ASL.

2 Preliminaries

2.1 A toy model

I Assumptions:

i Value is temporally separable.

ii Ex ante value is expected value.

iii We can assess changes in value relative to a baseline (status quo, inaction, . . .)

II Notation:

i V_o, L_o, S_o are overall, short-term and long-term value of an option o.

ii Δ relativizes value to a baseline.[1]

III Modeling the value of options:

$$E[\Delta V_o] = E[\Delta S_o] + E[\Delta L_o].$$

2.2 Swamping and convergent ASL

I Let a **swamping longtermist option** be one whose long-term value exceeds (in magnitude) the best-possible short-term value.

$$E[\Delta L_o] > max_{o' \in O} |E[\Delta S_{o'}]| .$$

II Two ways for ASL to hold:

[1] I.e. $\Delta(x) = x - V(b)$ where b is the baseline option.

i **Swamping ASL:** In a wide class of decision situations, the option that is *ex ante* best is (a) contained in a fairly small subset of options whose *ex ante* effects on the very long-run future are best, **and (b) a swamping longtermist option**.

ii **Convergent ASL:** ASL $\wedge \neg$ (Swamping ASL).

III **My target:** Swamping ASL. Reasons:

i Most revisionary.

ii Often what longtermists have in mind.

iii Drives widest argument from axiological to deontic strong longtermism.

2.3 The rarity thesis

I I aim to establish:

> **(Rarity thesis)** The vast majority of options we confront as decisionmakers are not swamping longtermist options.

3 The good case: Present-day, cause-neutral philanthropy

I Let a **strong swamping** option be one whose long-term effects *vastly* exceed (in magnitude) the best-achievable short-term effects, say:

$$E[\Delta L_o] > 10 * max_{o' \in O} |E[\Delta S_{o'}]| .$$

II **Argument from strong swamping:** In a decision problem $\mathcal{D}$:

i **Existence claim:** Some option in $\mathcal{D}$ is a strong swamping option.

ii **Existence → swamping ASL:** Any problem containing a strong swamping option satisfies swamping ASL.[2]

iii **Conclusion:** Therefore $\mathcal{D}$ satisfies swamping ASL.

III **Argument from single-track dominance:** That some option o in $\mathcal{D}$ is a strong swamping option.

> **(Single-track dominance)** The lion's share of the option's expected impact on the long-term future is driven by a single causal pathway or effect.

[2]Sketch argument: Let o be a strong swamping option. Note (a) that the best option in $\mathcal{D}$ is a swamping option, since a strong swamping option is better than any non-swamping option. Note (b) that the best option must have at least 9/10 of the best-achievable long-term value, since the existence of a strong swamping option implies that short-term value differences cannot make up any larger gap than this. Hence the best option is a swamping option with at least 9/10 of the best-achievable long-term value, and swamping ASL follows. (A technicality: depending on what you mean by a 'small' subset of near-best options, you might need additional conditions).

(Nontrivial probability of significant benefit) For some extremely large value of N, $Pr(\Delta L_o > N)$ is nontrivial, considering only effects along the specified track.[3]

(Relatively trivial probability of significant harm) With N as before, $Pr(\Delta L_o < -N) << Pr(\Delta L_0 > N)$, considering only effects along the specified track.

Note: Most options satisfying these premises will be strong swamping options.[4]

IV **An example:** Asteroid detection and the SpaceGuard survey:

i NASA classifies asteroids with diameter > 1km as catastrophic.

ii **Risks:**

i 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).

ii In 2019, an asteroid 100m in diameter passed five-times closer to earth than the average orbital distance of the moon. It was detected the day before.

iii Base probability of catastrophic impacts: 1/6,000 centuries (Stokes et al. 2017).

iii **Asteroid detection (past):** SpaceGuard survey mapped 95% of asteroids with diameter > 1km for $70m.

iv **Argument from single-track dominance:** Check premises individually.

v **More concretely:**

i **Estimated future humanlike lives:** 10^{13} to 10^{55} (Bostrom 2014; Newberry 2021).

ii **Stipulated probability that early warning would prevent extinction:** $1/10^6$.

iii **Expected cost per life saved:** Counting only impacts within a century: $7 on worst estimate. Peanuts on most.

iv **Best expected short-term cost of saving a life:** $2k-3k (GiveWell 2021).

4 First argument for rarity: Rapid diminution ('Thin tails')

I Consider the event partition $\{[\Delta L_o = k] : k \in \mathbb{R}\}$.

II Plausibly, as k becomes larger $Pr(\Delta L_o = k)$ decreases.

III **Rapid diminution** occurs when $Pr(\Delta L_o = k)$ diminishes faster than k increases.

[3] I.e. for specified N large and k nontrivial, $Pr(\Delta L_o > N) > k$, ignoring off-track effects.

[4] Strictly speaking, it is possible for an option with these features to fail to be a swamping longtermist option, for example if for some $M > N$ we have $Pr(\Delta L_o > M) < Pr(\Delta L_o < -M)$. I hope it is clear how the requirements might be tightened to avoid this worry, but also that this tightening would be a distraction in most cases of interest, where this is not a live worry.

IV **Consequence of rapid diminution:** $E[\Delta V_o]$ sees at most a modest impact from large far-future benefits.

V **Example of rapid diminution:** Normal distribution.

VI **Argument from rapid diminution:** We should expect many of the options we face to exhibit rapid diminution.

VII **The case for rapid diminution:** Given by **persistence skepticism**: many options do not make a large persisting impact on the long-term future.

VIII **Evidence from persistence studies: Notable failures**

- i **Might expect:** American bombing in Vietnam and Japan had persistent demographic and economic effects in hardest-hit areas.
- ii **Finding:** Effects on population size, poverty rates, and consumption patterns already below significance (Davis and Weinstein 2008; Miguel and Roland 2011).

IX **Evidence from persistence studies: Possible successes**

- i Introduction of plough → increased gendered division of labor (Alesina et al. 2011, 2013).
- ii African slave trade → present-day reductions in social trust and economic indicators (Nunn 2008; Nunn and Wantchekon 2011).
- iii Catholic Church → WEIRD personality traits.

X **Some grains of salt:**

- i These findings are controversial (Kelly 2019; Sevilla 2021).
- ii There aren't so many positive findings.
- iii (*) These examples involve comparably-large short-term effects.

XI **At the same time:** Clearly *some* events should be expected to have a lasting impact (example: Roman sacking of Carthage).

5 Second argument for rarity: Washing out ('Significant symmetry')

I **Washing out** occurs to the degree that potential positive and negative far-future impacts cancel one another in taking expectations.

II **First presentation of argument from washing out:** Evidential paucity.

- i **Symmetric priors:** Total ignorance should be represented by symmetric priors.
- ii **Evidential paucity:** We often have little evidence bearing on long-term effects.
- iii **Intermediate conclusion:** Posteriors over ΔL_o should show strong effects of priors.

iv **Final conclusion:** Posteriors should be heavily symmetric.

III **Second presentation of washing out:** forecasting pessimism.

i **Forecasting pessimism:** It's very difficult to reliably forecast the long-term future.

ii **Normative premise:** Uncertainty about forecasts should lead us to discount large estimates of far-future impact.

iii **(One possible) link to washing out:** Cancellation between potential optimistic and pessimistic forecasts.

IV **Arguments for forecasting pessimism:**

i **Limited track records:** Mixed and limited track-records at 20-30 years.[5] Almost nonexistent track-records at 100^+ years.

ii **Practitioner skepticism:** Many practitioners refuse to make predictions on this timescale, and recommend non-forecasting methods.

iii **Multidimensionality of value:** Predicting a single quantity (i.e. malaria deaths prevented) is easy. But many quantities determine overall value. So predicting V_o is hard.

V **Summing up the case for rarity:**

i By rapid diminution, we should think many options are unlikely to have large far-future impacts.

ii By washing out, we should often expect significant cancellation between potential positive and negative future impacts.

6 The good case revisited

I **The good news:** All of these arguments fall flat in the good case.

II **Theoretical argument:** Cross-apply premises from the argument from single-track dominance.

III **Concerete argument:** Check cases. (**Focus:** Asteroid detection).

i **Persistence skepticism:** Not being extinct is potentially quite persistent.

ii **Forecasting pessimism:**

i **Track records:** We have an *excellent* track record of long-term astronomical predictions.

ii **Practitioner skepticism:** Forecasts were from NASA.

iii **Multidimensionality of value:** Enough to believe future will be good.

[5]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).

7 First challenge to the scope of longtermism: The area challenge

I **Area challenge:** As we look at decision problems restricted to specific cause areas, swamping ASL often becomes less plausible.

II **Example:** Choice between two deworming programs.

III **Argument from single-track dominance fails:**

- i **Single-track dominance:** Implausible.
- ii **Nontrivial probability of significant benefit:** Hard to get this without denying (Relatively Trivial Probability of Significant Harm).

IV **Arguments for rarity get bite:**

- i **Forecasting pessimism:**
  - i **Track records:** No track record of predicting value of public health interventions at this timescale.
  - ii **Practitioner skepticism:** Even GiveWell doesn't try.
  - iii **Multidimensionality of value:** Paradigm case (Deaton 2015).
- ii **Persistence skepticism:** Hard to see what the argument could be other than a general attack on persistence skepticism.

8 Second challenge to the scope of longtermism: Option unawareness

I **A motivating example:**

> **(Not quite 007):** You're being chased down an alleyway by masked assailants. 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. And all of this could be inferred with high probability from your knowledge of physics together with your present perceptual evidence. But you haven't even considered it.

II **Unconstrained option set:** {turn right, turn left, fight, break pipe}.

III **Awareness-constrained option set:** {turn right, turn left, fight}.

IV **The case for awareness-constrained readings:** The ex-ante perspective involves limited awareness of options, just like it involves limited information.

V **Option unawareness in the good case:** No threat because:

- i We're already aware of swamping longtermist options.

ii If we weren't, we could fund *meta-options* like option generation research.

VI **Option unawareness beyond the good case:**

i By rarity, most options aren't swamping longtermist options.

ii By option unawareness, most choice sets are modestly sized.

iii Hence most choice sets don't contain any swamping longtermist options.

VII **An example: Global health**

i **Awareness-unconstrained problem:** Of *any possible* deworming intervention, which should I take?

ii **Awareness-constrained problem:** Of the leading international deworming programs, which should I fund?

VIII **Option generation is no help:** Often too costly!

9 Summing up

I **Swamping ASL:** In a wide class of decision situations, the option that is *ex ante* best is (a) contained in a fairly small subset of options whose *ex ante* effects on the very long-run future are best, **and (b) a swamping longtermist option**.

II **The good case:** Swamping ASL holds in contemporary cause-neutral philanthropy and adjacent decisions.

III **The rarity thesis:** The vast majority of options we confront as decisionmakers are not swamping longtermist options. **Motivations:**

i **Rapid diminution:** By persistence skepticism.

ii **Washing out:** By evidential paucity and forecasting skepticism.

IV **Two scope challenges:**

i **Area challenge:** The case for swamping ASL becomes less strong in many specific cause areas.

ii **Challenge from option unawareness:** Swamping ASL often fails in awareness-constrained decision problems.

References

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

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.

Bostrom, Nick. 2014. *Superintelligence*. Oxford University Press.

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

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.

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.

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

Greaves, Hilary and MacAskill, William. ms. "The case for strong longtermism." Unpublished manuscript.

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.

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

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

Newberry, Toby. 2021. "How many lives does the future hold?" Technical report, Global Priorities Institute.

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

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

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

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.

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

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.

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