Nick Beckstead - Global Priority Setting and Existential Risk: Crucial Ethical Considerations

Abstract

Several potential threats--such as catastrophic climate change, and the proliferation of nuclear and biological weapons, and the development of dangerous future technologies--could destroy human civilization in the next century. This raises an urgent question: just how important is it to decrease the chance of our extinction, relative to other good things that governments, private foundations, wealthy philanthropists, and private individuals could do with their time and resources?

The answer is highly sensitive to the following ethical questions:

1. Is it important that people exist in the future?
2. If so, does humanity's flourishing in additional future periods have diminishing marginal value?
3. Should we discount future benefits?
4. Should we maximize expected value in one-shot, low-probability, high-stakes gambles?

In the first part of my dissertation, I defend the answers: Yes, No, No, and Yes (a response characteristic of the risk-neutral aggregative consequentialist tradition). These responses strongly suggest that decreasing existential risk is probably the most urgent global priority, even more important than funding the world's most cost-effective developing world health interventions, which might reasonably be regarded as the best alternative use of resources.

However, as I later show, these responses raise significant problems. For almost any evidential state, they imply that we should devote all of our resources to pursuing infinitely valuable outcomes, given any non-zero probability of success. This conclusion is very hard to accept, so I argue that we should insist on using a bounded utility function. Insisting on a bounded utility function has some implausible consequences which we must come to terms with, including implausibly risk-seeking behavior in very bad cases and sensitivity to seemingly irrelevant details about the distant past. This approach makes it less clear, but still plausible, that decreasing existential risk should be the dominant global priority.

Proposal

Chapter 1: How Could We Be So Wrong?

My argumentative style stresses distrust of intuitions, a desire to explain away recalcitrant intuitions in terms of known biases, a high premium on simplicity, and a preference for fitting theories to wide classes of moral judgments, rather than putting a few judgments under a microscope. I argue that this is the rational response to three sources of evidence indicate that our intuitive ethical judgments are less reliable than we might have hoped: a historical record of accepting morally absurd social practices; a scientific record showing that our intuitive judgments are systematically governed by a host of heuristics, biases, and irrelevant factors; and a philosophical record showing deep, probably unresolvable, inconsistencies in common moral convictions. These methodological positions inform the rest of the dissertation, and help explain how people could have ignored the enormous ethical weight behind the interests of future generations.

Chapter 2: The Case for Focusing on Existential Risk

Next, in a chapter that structures much of the dissertation, I explain how the importance of mitigating existential risks is highly sensitive to the four questions discussed above, and why the answers I identified are, prima facie, plausible. If we answer as indicated above (in line with risk-neutral aggregation), decreasing existential risk becomes the top global priority. If these conditions are significantly relaxed, what happens in the far future matters much less, and other priorities, such as developing world health, become more pressing.

Chapter 3: Should "Extra" People Count for Less?

We might believe that it does not matter whether there are any future people at all, except insofar as their existence affects the present generation. The philosophical rationale for this is roughly that if these people never exist, they are not harmed by their non-existence. If we also are not harmed or benefited by their non-existence, then their non-existence would matter to no one, and therefore would not matter at all. I argue that, though influential, this position has implausible implications about the permissibility of having children and the desirability of avoiding premature extinction, and must be rejected.

Chapter 4: Do Additional Generations Have Diminishing Marginal Value?

My argument for taking existential risks much more seriously could also be resisted by claiming that once enough generations have existed, it matters less whether additional generations exist. In this chapter, I defend the claim that the importance of the existence of a particular generation is independent of what happened in previous generations. Competing views implausibly imply that how good it would be to avert a catastrophe depends on events which might have occurred in the distant past. Moreover, if this view is strong enough to dismantle the case for focusing on existential risk, it would have additional implausible consequences. In particular, it would imply that if civilization continued for sufficiently long, it would be essentially irrelevant whether it came to a premature end.

Chapter 5: What Kinds of Partiality to Nearer Generations Could Be Justified?

This chapter explores an alternative justification for caring more about generations that are nearer in time: we are more emotionally connected to them and they are more likely to develop our culture and ideals, whereas generations in the distant future will be less related to us and much different from us. It might be argued that this is the true ethical relevance of ensuring the existence of future generations, rather than the idea that their existence and flourishing is valuable in itself.

I argue that though we may justifiably be somewhat more concerned about the persistence of the next couple of generations, this concern cannot, by itself, explain the value of future generations. The explanation on offer cannot explain how people could rationally value the existence of people who are unrelated to them, or take comfort in the fact that even if life on earth does not survive, there may be intelligent life elsewhere in the Universe.

Chapter 6: Recklessness, Timidity, and Fanaticism: The Rationality of Longshots

In this chapter, I show that the answers to the four questions defended in previous chapters imply a very implausible conclusion: in almost any evidential state, we should expend all of our resources pursuing infinitely good outcomes (or trying to avoid infinitely bad ones). I argue that this problem is best avoided by adopting a bounded utility function. Adopting a bounded utility function has some significant costs, which I enumerate, but it is preferable to the alternative.

Chapter 7: Conclusion

This chapter will synthesize some conclusions from previous chapters and attend to the question of how to resolve remaining uncertainty about the ethical assumptions underpinning the evaluation of existential risks. I argue that given the adoption of a bounded utility function, it is still plausible, though much less obvious, that reducing existential risk should be the top global priority.