A RESEARCH AGENDA FOR THE

GLOBAL PRIORITIES INSTITUTE

Global Priorities Institute

Version 2.1
October 2020

Table of contents

Introduction

The Global Priorities Institute's vision and mission

There are many problems in the world. Because resources are scarce, it is impossible to solve them all. An actor seeking to improve the world as much as possible therefore needs to prioritise, both among the problems themselves and among means for tackling them.

This task of prioritisation requires careful analysis. Some opportunities to do good are vastly more cost-effective than others. But identifying which are the better opportunities requires grappling with a host of complex questions: questions about how to evaluate different outcomes, how to predict the effects of actions, how to act in the face of uncertainty, how to identify more practically usable proxies for social improvements, and many other topics.

In practice, only a minority of actors (whether individual or institutional) regularly give significant explicit consideration to the question of which option would do the most good, considered impartially. There are many reasons for this. Some concern constraints imposed by politics, or other limits of motivation. One significant stumbling block is that *there is simply not enough information or understanding about what it would look like* to determine priorities and actions on the basis of an intellectually rigorous, evidence-based assessment of the amount of good that candidate options are expected to achieve, all things considered, in the long run, and in impartial terms. This issue should be of importance to the adherents of many different moral views, including non-consequentialists, who standardly assign some significance to promoting the good.

A significant exception to this general tendency of decision makers to neglect the impartial good when setting priorities is found in the effective altruism community. Over the past ten years or so, this growing community has devoted a rapidly increasing flow of resources, both intellectual and financial, to the enterprise of promoting morally valuable outcomes as effectively as possible. For example, the [Open Philanthropy Project](#) has made approximately 1,000 philanthropic grants with a total worth of more than $1.1bn since 2012, and [80,000 Hours has tracked thousands of people](#) who have made significant changes to their career plans based on its research and recommendations. The movement has developed numerous novel and exciting ideas, and has been audacious in pushing forward the implementation of those ideas. However, many of these ideas have yet to be explored in academia.

The Global Priorities Institute (GPI) exists to develop and promote rigorous, scientific approaches to the question of how appropriately motivated actors can do good more effectively. Our core belief is that the existence of a wide base of high-quality research on these questions, and (relatedly) an increased focus on those questions within academia, is a prerequisite for the widespread adoption of an effectiveness-based approach to global prioritisation.

This line of thought motivates the following vision and mission:

GPI's vision

A world in which global priorities are set by using evidence and reason to determine what will do the most good.

GPI's mission

To conduct and promote world-class, foundational academic research on how most effectively to do good.

GPI's research agenda

The central focus of GPI is what we call 'global priorities research': research into issues that arise in response to the question, 'What should we do with a given amount of limited resources if our aim is to do the most good?' This question naturally draws upon central themes in the fields of economics and philosophy.

Thus defined, global priorities research is in principle a broad umbrella. Within that umbrella, this research agenda sets out the research themes that GPI is particularly interested in at the present time. **We stress that in all cases, what we list here are relatively broad research themes, rather than the more specific research questions that would naturally correspond to individual research papers.** Within each such theme, the first step is to do significant further work identifying and articulating the fruitful research questions. The document is structured as follows.

Section 1 outlines what we call *the longtermism paradigm*. This paradigm centres around the idea that because of the potential vastness of the future portion of the history of humanity and other sentient life, the primary determinant of optimal policy, philanthropy and other actions is the effects of those policies and actions on the very long-run future, rather than on more immediate considerations. Because these ideas seem plausible, seem likely to have fairly radically revisionary implications if correct, and are currently underexplored in academic research, **this is the main focus of GPI's own research** (at the time of writing and, we predict, for at least the next one to two years). **We are particularly keen to hear from other researchers who share this interest.**

Section 2 concerns general issues in cause prioritisation. This section covers issues that are not specific to a longtermist point of view, but that arise for agents engaged in global prioritisation.

Appendix A indicates additional areas of possible research that would further GPI's mission, but that GPI itself is not working on now or for the immediately foreseeable future, for reasons of capacity and focus. Appendix B indicates areas of existing academic literature that serve as particularly relevant background for the topics on this research agenda. Appendix C contains links to additional informal discussion of the themes covered in this research agenda.

The intended audience for this document is academics (especially, but not only, in economics and philosophy) who are potentially interested in working with GPI, whether as GPI researchers or as external collaborators, or who are otherwise interested in the same mission.

1. The longtermism paradigm

As noted above, an actor seeking to improve the world as much as possible with limited resources needs to prioritise: which problems should she focus on and which steps should she take to address those problems, to the exclusion of others?

Key to GPI's approach to this question is what we call *the longtermism paradigm*. This paradigm has two key components. First, insofar as the social value of policies and actions is determined by their consequences or effects, all effects should be taken into account, and not only those that are in any specified sense 'direct'. Second, the consequences or effects of policies and actions should be given the same weight regardless of where or when in space and time they occur. Importantly, this implies that evaluations of policies and actions should have a zero rate of pure time preference and should not assign different weights to individuals' welfare based on geographical location.

This paradigm has potentially radical implications. Given how long human and other sentient life could potentially survive, it suggests that the primary determinant of differences in social value among the best actions and policies that could be pursued today could well be their effects on the very long-term future, rather than any effects within our own lifetimes. In contrast, mainstream economics and policy research typically takes the perspective that improving the course of the far future is not tractable and is relatively unimportant under frameworks that use a positive discount rate.

This contrast warrants further research on the articulation, evaluation, implications and implementation of longtermist ideas in global prioritisation.

1.1. Articulation and evaluation of longtermism

Let us define *longtermism* as the view that the primary determinant of the differences in social value among actions and policies available today is the effect of those actions and policies on the very long-term future. This view is supported by plausible arguments, and has widespread significance if correct. This warrants more research to articulate, evaluate and explore the implications of a longtermist view.

Philosophy:

- It is natural to think that in evaluating interventions, we should take into account *all* welfare-relevant effects of those interventions, not only those that are intended or direct. The argument that we should value indirect effects seems in tension with the view, held by some philosophers, that when deciding whom to aid, we are generally morally constrained to consider only the direct impact of our actions on those we can help, as opposed to the indirect utility of helping some rather than others (Kamm 1993; Brock 2003; Lippert-Rasmussen and Lauridsen 2010; Du Toit and Millum 2016). How is this tension best resolved (Mogensen 2020a) (INFORMAL: Greaves 2015)?

- Assuming that indirect effects should be counted, that future welfare should not be discounted, and that suitable population axiological assumptions are adopted, provide a rigorous articulation of the case for thinking that the primary determinant of differences in social value between the best policies and actions available today is the expected effects of those policies and actions on the very far future (Bostrom 2003) (INFORMAL: Karnofsky 2014; Todd 2017). How sensitive is this argument to variations in evaluative assumptions over which there is reasonable disagreement (Beckstead 2013; Beckstead forthcoming) (INFORMAL: Ord 2017; Sittler 2018)?

PHIL - ETHICS OF DISCOUNTING

- Any actions and policies that affect the very long-term future will change the size and composition of the total population of everyone who will ever live. The moral evaluation of such changes is fraught with paradoxes (Parfit 1984; Ng 1989; Arrhenius MS). Does longtermism presuppose some particular, controversial population axiology, such as total utilitarianism? Or might longtermism be robustly supported across a range of minimally plausible population axiologies? If so, do different axiologies support different conclusions about intra-longtermist prioritisations (Beckstead 2013; Greaves and MacAskill 2019; Thomas 2019; Mogensen 2020b; Tarsney and Thomas 2020)?

PHIL - POPULATION ETHICS

Economics:

- Under what conditions would a social planner or philanthropist prioritise policies that primarily increase social welfare in the far future rather than in the near term? For instance, under what condition would such agents prioritise saving for future generations (Ramsey 1928) or reducing the risk of human extinction (Baranzini and Bourguignon 1995)?

ECON - DISCOUNTING, OPTIMAL SAVING, CATASTROPHIC RISKS

- What are the long-term effects of interventions that seem particularly high-priority from a short-term perspective (Athey et al. 2019; Bouguen et al. 2018), for example the highest impact global health and development interventions (INFORMAL: Karnofsky 2013) or the most effective programmes to help farm animals (Matheny and Chan 2005) (INFORMAL: Shulman 2013)? Under what conditions, if any, might they exceed the expected long-term impacts of (other) efforts aimed explicitly at improving the long term?

ECON - FORECASTING, PROGRAMME EVALUATION

Interdisciplinary:

- There is already a substantial literature (on both sides) evaluating the claim that public policy should adopt a zero rate of pure time preference (Greaves 2017; Dasgupta 2008). Given the importance of this claim to the longtermism paradigm, further research that changes the balance of arguments on this question could still be valuable. What more, if anything, can be said on the matter (Mogensen 2019)?

PHIL - ETHICS OF DISCOUNTINGECON - DISCOUNTING

- To what extent do considerations of saturation (e.g., the possibility that utility as a function of consumption is bounded) constrain the possibilities for leveraging the

vast potential size of the future to identify policies and actions with extremely high social value (Ng 1991)?

Existing academic literature:

- Arrhenius, Gustaf. *Population Ethics: the Challenge of Future Generations*. Unpublished manuscript.

- Athey, Susan, Raj Chetty, Guido Imbens and Hyunseung Kang. 'The Surrogate Index: Combining Short-term Proxies to Estimate Long-term Treatment Effects More Rapidly and Precisely'. *Randomized Social Experiments Ejournal*, 2003.

- Baranzini, Andrea and François Bourguignon. 'Is Sustainable Growth Optimal?' *International Tax and Public Finance* 2.2 (1995): 341–356.

- Beckstead, Nicholas. 'A Brief Argument for the Overwhelming Importance of Shaping the Far Future'. In *Effective Altruism: Philosophical Issues*, edited by Theron Pummer and Hilary Greaves. Oxford: Oxford University Press, 2019.

- ———. 'On the Overwhelming Importance of Shaping the Far Future'. PhD dissertation. New Brunswick: Rutgers University, 2013.

- Bouguen, Adrien, Yue Huang, Michael Kremer and Edward Miguel. 'Using RCTs to Estimate Long-Run Impacts in Development Economics'. No. w25356. National Bureau of Economic Research, 2018.

- Bostrom, Nick. 'Astronomical Waste: The Opportunity Cost of Delayed Technological Development'. *Utilitas* 15, no. 3 (2003): 308–314.

- Brock, Dan W. 'Separate Spheres and Indirect Benefits'. *Cost Effectiveness and Resource Allocation* 1, no. 4 (2003).

- Dasgupta, Partha. 'Discounting climate change.' *Journal of Risk and Uncertainty* 37.2–3 (2008): 141–169.

- Greaves, Hilary. 'Discounting for Public Policy: A Survey'. *Economics & Philosophy* 33, no. 3 (2017): 391–439.

- ———. 'Discounting Future Health'. In *Global Health Priority-Setting: Cost-Effectiveness and Beyond*, edited by Ruger and Verguet Otterson Millum Johansson Jamison Emanuel Norheim. Oxford: Oxford University Press, forthcoming.

- Kamm, Frances M. *Morality, Mortality; Volume I: Death and Whom to Save From It*. Oxford: Oxford University Press, 1998.

- Lippert-Rasmussen, Kasper and Sigurd Lauridsen. 'Justice and the Allocation of Healthcare Resources: Should Indirect, Non-Health Effects Count?' *Medicine, Health Care and Philosophy* 13, no. 3 (2010): 237–246.

- Matheny, Gaverick and Kai Chan. 'Human Diets and Animal Welfare: The Illogic of the Larder'. *Journal of Agricultural and Environmental Ethics* 18, no. 6 (2005): 579–594.

- Mogensen, Andreas L. '"The only ethical argument for positive δ"?' Working paper, 2019.

- ———. 'Meaning, Medicine, and Merit'. *Utilitas* 32, no.1 (2020a), 90–107.

- ———. 'Moral demands and the far future'. *Philosophy and Phenomenological Research*, 2020b.

- Ng, Yew-Kwang, 'What Should We Do About Future Generations? Impossibility of Parfit's Theory X'. *Economics and Philosophy* 5, no. 2 (1989): 235–253.

- ———. 'Should we be very cautious or extremely cautious on measures that may involve our destruction?.' *Social Choice and Welfare* 8, no. 1 (1991): 79–88.

- Parfit, Derek. *Reasons and Persons*. Oxford: Oxford University Press, 1984.

- Ramsey, Frank Plumpton. 'A Mathematical Theory of Saving.' *The Economic Journal* 38.152 (1928): 543–559.

- Tarsney, Christian and Teruji Thomas. 'Non-Additive Axiologies in Large Worlds'. Working paper, 2020.

- Thomas, Teruji. 'The Asymmetry, Uncertainty, and the Long-Term'. Working paper, 2019.

- Toit, Jessica du and Franklin Miller. 'The Ethics of Continued Life-Sustaining Treatment for Those Diagnosed as Brain-Dead'. *Bioethics* 30, no. 3 (2016): 151–158.

Existing informal discussion:

- Nick Bostrom, Crucial considerations and wise philanthropy, 2014.

- Hilary Greaves, Repugnant interventions, 2015.

- Holden Karnofsky, Flow-through effects, 2013.

- Holden Karnofsky, The Moral Value of the Far Future, 2014.

- Toby Ord, Why the long-term future of humanity matters more than anything else, and what we should do about it, 2017.

- Carl Shulman, Vegan advocacy and pessimism about wild animal welfare, 2013.

- Thomas Sittler, The expected value of the long-term future, 2018.

- Benjamin Todd, Introducing longtermism, 2017a.

- Benjamin Todd, The case for reducing extinction risk, 2017b.

- Brian Tomasik, Charity Cost-Effectiveness in an Uncertain World, 2013.

1.2. The value of the future of humanity

Longtermism is often thought to lead to the conclusion that extinction risk reduction should be a global priority. This presupposes that the expected social value of continued human existence is positive. But one can imagine scenarios and social welfare criteria according to which humanity's future should be expected to contain more bad than good. Before engaging in more fine-grained cause prioritisation across efforts to reduce extinction

risk, it is therefore important to consider the sign and magnitude of the expected social value of the continued existence of humanity.

Philosophy:

- Assess the expected value of the continued existence of humanity. Might this expected value be negative, or just unclear (INFORMAL: Christiano 2013; West 2017)? How do our answers to these questions vary if we (i) assume utilitarianism (INFORMAL: Shulman 2012; Dickens 2015); (ii) assume a non-utilitarian axiology (INFORMAL: Greaves, 2016; Brauner and Grosse-Holz 2018); (iii) fully take axiological uncertainty into account (Greaves and Ord 2017; MacAskill 2019)?

PHIL - MORAL UNCERTAINTY, POPULATION ETHICS

- Should longtermists be more concerned about avoiding the worst possible outcomes for the future than ensuring that the very best outcomes occur? This could be because the costs of the worst outcomes are greater than the benefits of the best outcomes, because avoidance of the worst outcomes is more neglected, or because the worst outcomes should be given more weight than the best outcomes (Hurka 2010) (INFORMAL: Althaus and Gloor 2018; Tomasik 2018)? If so, what activities would be best? (MacAskill MS) (INFORMAL: Gloor 2018)

PHIL - MORAL UNCERTAINTY, DECISION THEORY

Economics:

- In analogy to the literature on the value of a statistical life, economists researching catastrophe have introduced the concept of the 'value of a statistical civilisation', which may be interpreted as the cost of premature human extinction (Weitzman 2009; Nordhaus 2011; Bommier et al. 2015; Fleurbaey et al. 2019). How should the value of statistical civilisation be estimated and what is its sign and magnitude?

ECON - CATASTROPHIC RISK, WELFARE ECONOMICS

Interdisciplinary:

- Social welfare criteria that are used to compare states that differ in population size typically specify a critical welfare level at which lives that are added to the population have zero contributive value to social welfare (Blackorby et al. 1995; Ng 1986; Broome 2004; Ch. 10). What kinds of lives have zero contributive value in this sense (Cockburn et al. 2014; Cowie 2017)?

PHIL - POPULATION ETHICS   ECON - WELFARE ECONOMICS

- Does the idea of option value provide strong reason to prevent human extinction even if there is uncertainty about the social value of the future (MacAskill MS)? What is the chance that future decision-makers will use humanity's 'cosmic endowment' such that current generations would be willing, now, to defer to them?

PHIL - MORAL UNCERTAINTY   ECON - VALUE OF INFORMATION, INTERGENERATIONAL GOVERNANCE

Existing academic literature:

- Blackorby, Charles, Walter Bossert and David Donaldson. 'Intertemporal population ethics: critical-level utilitarian principles'. *Econometrica: Journal of the Econometric Society* (1995): 1,303–1,320.

- Bommier, Antoine, Bruno Lanz and Stéphane Zuber. 'Models-as-usual for unusual risks? On the value of catastrophic climate change'. *Journal of Environmental Economics and Management* 74 (2015): 1–22.

- Broome, John. *Weighing Lives*. Oxford; New York: Oxford University Press, 2004.

- Cockburn, John, Jean-Yves Duclos, and Agnès Zabsonré. 'Is global social welfare increasing? A critical-level enquiry'. *Journal of Public Economics* 118 (2014): 151–162.

- Cowie, Christopher. 'Does the repugnant conclusion have any probative force?'*Philosophical Studies* 174.12 (2017): 3,021–3,039.

- Fleurbaey, Marc, Maddalena Ferranna, Mark Budolfson, Francis Dennig, Kian Mintz-Woo, Robert Socolow, Dean Spears and Stéphane Zuber. 'The Social Cost of Carbon: Valuing Inequality, Risk, and Population for Climate Policy'. *The Monist* 102.1 (2019): 84-109.

- Greaves, Hilary. 'Optimum Population Size'. In *Oxford Handbook of Population Ethics*, edited by Gustaf Arrhenius, Krister Bykvist and Tim Campbell. Oxford: Oxford University Press, forthcoming.

- Greaves, Hilary and Toby Ord. 'Moral Uncertainty About Population Axiology'. *Journal of Ethics & Social Philosophy* 12, no. 2 (2017): 135–167.

- Hurka, Thomas. 'Asymmetries In Value'. *Noûs* 44, no. 2 (2010): 199–223.

- MacAskill, William. 'Human Extinction, Asymmetry, and Option Value'. Manuscript in preparation.

- ———. 'Practical Ethics Given Moral Uncertainty'. *Utilitas* 31, no. 3 (2019): 231–245.

- MacAskill, William, Krister Bykvist and Toby Ord. *Moral Uncertainty*. Oxford University Press, 2020.

- Ng, Yew-Kwang. 'Social criteria for evaluating population change: an alternative to the Blackorby-Donaldson criterion'. *Journal of Public Economics* 29.3 (1986): 375–381.

- Nordhaus, William D. 'The economics of tail events with an application to climate change'. *Review of Environmental Economics and Policy* 5.2 (2011): 240–257.

- Weitzman, Martin L. 'On Modeling and Interpreting the Economics of Catastrophic Climate Change'. *Review of Economics and Statistics* 91, no. 1 (2009): 1–19.

Existing informal discussion:

- David Althaus and Lukas Gloor, Reducing Risks of Astronomical Suffering: a Neglected Priority, 2018.

- Jan Brauner and Friederike Grosse-Holz, [The expected value of extinction risk reduction is positive](), 2018.

- Paul Christiano, [Why might the future be good?](), 2013.

- Michael Dickens, [Is Preventing Human Extinction Good?](), 2015.

- Lukas Gloor, [Cause prioritization for downside-focused value systems](), 2018.

- Hilary Greaves, [Extinction risk and population ethics](), 2016.

- Carl Shulman, [Spreading happiness to the stars seems little harder than just spreading](), 2012.

- Carl Shulman, [Are pain and pleasure equally energy-efficient?](), 2012.

- Brian Tomasik, [Risks of Astronomical Future Suffering](), 2018.

- Ben West, [An Argument for Why the Future May Be Good](), 2017.

1.3. Reducing and mitigating catastrophic risk

Human civilisation arguably has the capacity to produce vast amounts of value over the course of the future (Ord 2020). If this is correct, then it may be uniquely important from a longtermist perspective to minimise the risk of catastrophes, such as near-term human extinction. The precise implications of this argument, however, warrant further scrutiny.

Philosophy:

- What do the most plausible person-affecting views in population ethics say about the value of reducing extinction risk (Thomas 2019) (INFORMAL: Greaves 2016)?

PHIL - POPULATION ETHICS

Economics:

- When confronted with multiple catastrophic risks, what is the optimal allocation of funds between investments aimed at preventing catastrophes, those aimed at mitigating their consequences, and other social investments (Martin and Pindyck 2015)?

ECON - CATASTROPHIC RISK, GROWTH THEORY

- How can endogenous growth models be adapted to weigh the benefits that growth may pose for the long term against the catastrophic risks that may come with technological development (Jones 2016a; Aschenbrenner 2020)?

ECON - GROWTH, CATASTROPHIC RISK

- Most of the work in economics concerning long-term catastrophic risk mitigation has focused on climate change. To what extent does climate change pose a genuinely existential threat (Méjean et al. 2017; Ord 2020) (INFORMAL: Halstead, 2018)? How do efforts to mitigate the risks of climate change compare with those designed to mitigate more neglected catastrophic risks (Martin and Pindyck 2015, 2017; Ord 2020) (INFORMAL: Duda 2016)?

ECON - CATASTROPHIC RISK, ENVIRONMENTAL ECONOMICS

- The expected social value of reducing the risk of human extinction in the near future is particularly high if we live in a critical time period such that the extinction risk is expected to be both high in the coming decades (or centuries) and very low for most of the future conditional on humanity surviving the coming decades (Sittler 2018; Ord 2020). How will the risk of human extinction evolve over time? For instance, should we expect the risk to decrease as economic growth leads to increasing investments in safety (Jones 2016a; Aschenbrenner 2020) or as humanity develops the capability for space colonisation?

ECON - CATASTROPHIC RISK, GROWTH THEORY

Interdisciplinary:

- Should longtermist policymakers and philanthropists focus on reducing the risk of human extinction (Matheny 2007) (INFORMAL: Todd 2017)? Alternatively, should they focus on achieving trajectory changes or other ways of increasing the expected value of the far future conditional on the survival of humanity (Beckstead 2013; Ng 2016; Méjean et al. 2017)?

ECON - GROWTH, CATASTROPHIC RISK

- The expected social value of implementing policies to reduce catastrophic risk could be substantial even if those policies only had a very small chance of succeeding. Is expected utility maximisation the correct approach for dealing with decisions of this character (Bostrom 2009; Smith 2014; Isaacs 2016; Monton 2019; Tarsney 2019; Peterson 2020)? Does any plausible alternative decision theory imply that low-probability, high-payoff opportunities are not among the interventions with the highest social value (INFORMAL: Karnofsky 2011)?

PHIL - DECISION THEORY ECON - DECISION THEORY

Existing academic literature:

- Aschenbrenner, Leopold. 'Existential Risk and Growth'. Working paper, 2020.

- Beckstead, Nicholas. 'On the Overwhelming Importance of Shaping the Far Future'. PhD dissertation. New Brunswick: Rutgers University, 2013.

- Bostrom, Nick. 'Existential Risk Prevention as Global Priority'. *Global Policy* 4, no. 1 (2013): 15–31.

- ———. 'Existential Risks: Analyzing Human Extinction Scenarios and Related Hazards'. *Journal of Evolution and Technology* 9 (2002).

- ———. 'Pascal's Mugging'. *Analysis* 69, no. 3 (2009): 443–445.

- Cotton-Barratt, Owen and Toby Ord. 'Existential Risk and Existential Hope: Definitions'. *Future of Humanity Institute*, Technical Report #2015-1.

- Isaacs, Yoaav. 'Probabilities cannot be rationally neglected'. *Mind* 125, no 499 (2016) 759–762.

- Méjean, Aurélie, Antonin Pottier, Stéphane Zuber and Marc Fleurbaey. 'Intergenerational Equity under Catastrophic Climate Change'. Working paper, 2017.

- Jones, Charles I. 'Life and Growth'. *Journal of Political Economy* 124, no. 2 (2016): 539–578.

- Martin, Ian W. R. and Robert S. Pindyck. 'Averting Catastrophes: The Strange Economics of Scylla and Charybdis'. *American Economic Review* 105, no. 10 (2015): 2,947–2,985.

- ———. 'Averting Catastrophes that Kill'. Working paper, 2017.

- Matheny, Jason G. 'Reducing the Risk of Human Extinction'. *Risk Analysis* 27, no. 5 (2007): 1,335–1,344.

- Millner, Antony. 'On Welfare Frameworks and Catastrophic Climate Risks'. *Journal of Environmental Economics and Management* 65, no. 2 (2013): 310–325.

- Monton, Bradley. 'How to avoid maximizing expected utility'. *Philosophers' Imprint* 19 (2019): 1–25.

- Ng, Yew-Kwang. 'The Importance of Global Extinction in Climate Change Policy'. *Global Policy* 7, no. 3 (2016): 315–322.

- Ord, Toby. *The Precipice: Existential Risk and the Future of Humanity*. London: Bloomsbury, 2020.

- Peterson, Martin. 'The St. Petersburg Paradox'. In *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta, 2020.

- Smith, Nicholas JJ. 'Is evaluative compositionality a requirement of rationality?' *Mind* 123, no. 490 (2014): 457–502.

- Weitzman, Martin L. 'On Modeling and Interpreting the Economics of Catastrophic Climate Change'. *Review of Economics and Statistics* 91, no. 1 (2009): 1–19.

Existing informal discussion:

- Roman Duda, Climate change (extreme risks), 2016.

- Hilary Greaves, Extinction risk and population ethics, 2016.

- John Halstead, Is climate change an existential risk?, 2018.

- Holden Karnofsky, Why We Can't Take Expected Value Estimates Literally (Even When They're Unbiased), 2011.

- Toby Ord, The timing of labour aimed at reducing existential risk, 2014.

- Benjamin Todd, The case for reducing extinction risk, 2017.

1.4. Other ways of leveraging the size of the future

The 'size' of the future may present us with other ways, beyond reducing or mitigating catastrophic risks, of producing vast amounts of value. In particular, we may be able to produce lasting technological or civilisational 'trajectory changes' whose expected long-term value exceeds that of existential risk mitigation. This warrants putting thought into identifying promising trajectory-change opportunities and developing a framework for prioritising among them.

Philosophy:

- Let *finitism* be the claim that, even if we ought to aim to bring about an astronomically large finite amount of value in the future, we ought not to aim explicitly to bring about an infinitely large amount of value. Is finitism defensible? If it is not defensible, is this a *reductio* of the idea that we ought to try to bring about an astronomically large finite amount of value, or an argument that we really should be pursuing infinite amounts of value (Vallentyne and Kagan 1997; Basu and Mitra 2003; Vallentyne and Lauwers 2004; Zame 2007; Asheim 2010; Bostrom 2011; Arntzenius 2014) (INFORMAL: West 2015)?

PHIL - DECISION THEORY, INFINITE ETHICS

Interdisciplinary:

- What kinds of trajectory change or other interventions might offer opportunities with very high expected social value as a result of the potential vastness of the future? Can we construct a useful taxonomy for thinking about these (Bostrom 2005) (INFORMAL: Duda 2017; Duda 2018; Beckstead 2014; Whittlestone 2017; Baum et al. 2019)?

PHIL - POPULATION ETHICS   ECON - GROWTH

- For what kinds of actions and policies do we expect effects to 'wash out' over very long timescales rather than to persist? Are their long-run effects typically of much greater expected value (whether positive or negative) than their short-run effects, taking into account both the vastness of the future and the generally greater uncertainty of effects that are more causally remote (Beckstead 2013) (INFORMAL: Beckstead 2013)?

PHIL - ETHICS OF CHAOS   ECON - TIME SERIES ECONOMETRICS, ECONOMIC HISTORY AND DEVELOPMENT

- How should a social planner rank alternatives involving possibilities of infinite quantities of social value (Vallentyne and Kagan 1997; Basu and Mitra 2003; Vallentyne and Lauwers 2004; Zame 2007; Asheim 2010; Bostrom 2011; Arntzenius 2014) (INFORMAL: West 2015)?

PHIL - INFINITE ETHICS   ECON - INTERGENERATIONAL EQUITY

Existing academic literature:

- Asheim, Geir B. 'Intergenerational Equity'. *Annual Review of Economics* 2, no. 1 (2010): 197–222.

- Arntzenius, Frank. 'Utilitarianism, Decision Theory, and Eternity'. *Philosophical Perspectives* 28, no. 1 (2014): 31–58.

- Basu, Kaushik and Tapan Mitra. 'Aggregating Infinite Utility Streams with Intergenerational Equity: The Impossibility of Being Paretian'. *Econometrica* 71, no. 5 (2003): 1,557–1,563.

- Baum, Seth D., Stuart Armstrong, Timoteus Ekenstedt, Olle Häggström, Robin Hanson, Karin Kuhlemann, Matthijs M. Maas, James D. Miller, Markus Salmela, Anders Sandberg, Kaj Sotala, Phil Torres, Alexey Turchin and Roman V. Yampolskiy. 'Long-Term Trajectories of Human Civilization'. *Foresight*, 2019.

- Beckstead, Nicholas. *On the Overwhelming Importance of Shaping the Far Future*. PhD dissertation. New Brunswick: Rutgers University, 2013.

- Bostrom, Nick. 'Infinite Ethics'. *Analysis and Metaphysics*, no. 10 (2011): 9–59.

- ———. *Superintelligence: Paths, Dangers, Strategies*. First edition. Oxford: Oxford University Press, 2014.

- ———. 'Technological Revolutions: Ethics and Policy in the Dark'. In *Nanoscale*, edited by Nigel M. de S. Cameron and M. Ellen Mitchell. Hoboken, NJ: John Wiley & Sons, 2007: 129–152

- Lauwers, Luc and Peter Vallentyne. 'Infinite Utilitarianism: More Is Always Better'. *Economics & Philosophy* 20, no. 2 (2004): 307–330.

- Vallentyne, Peter and Shelly Kagan. 'Infinite Value and Finitely Additive Value Theory'. *The Journal of Philosophy* 94, no. 1 (1997): 5–26.

- Zame, William R. 'Can intergenerational equity be operationalized?' *Theoretical Economics* 2, no. 2 (2007): 187–202.

Existing informal discussion:

- Nick Beckstead, A Proposed Adjustment to the Astronomical Waste Argument, 2013.

- Nick Beckstead, A relatively atheoretical perspective on astronomical waste, 2014.

- Paul Christiano, Against moral advocacy, 2013.

- Roman Duda, Building effective altruism, 2017.

- Roman Duda, Global priorities research, 2018.

- Carl Shulman, Spreading happiness to the stars seems little harder than just spreading, 2012.

- Ben West, Problems and Solutions in Infinite Ethics, 2015.

- Jess Whittlestone, Improving institutional decision-making, 2017.

1.5. Intergenerational governance

Many of the long-term plans made by policymakers and philanthropists are vulnerable to being altered or undone by future generations. In evaluating the long-term consequences of policies and actions, it is therefore important to reckon with questions of how to influence the behaviour of future policymakers, and how to 'coordinate' optimally in the face of constraints on that influence.

Philosophy:

- The idea of the *long reflection* is that of a long period – perhaps tens of thousands of years – during which human civilisation, perhaps with the aid of improved cognitive ability, dedicates itself to working out what is ultimately of value (INFORMAL: MacAskill 2018). It may be argued that such a period would be warranted before deciding whether to undertake an irreversible decision of immense importance, such

as whether to attempt spreading to the stars. Do we find ourselves, or are we likely to find ourselves, in a situation where a long reflection would in fact be warranted? If so, how should it be implemented?

- How does the adoption of a long-term perspective that rejects a positive rate of pure time preference, shape debates about feasibility, idealisation, and utopianism in political theory (Sen 2009; Lawford-Smith 2013; Estlund 2019)? Over very long timescales, feasibility constraints in politics weaken. Does a long-term perspective therefore support a renewed role for utopian political theorising? Or does it argue against a focus on utopian blueprints, in favour of designing open, exploratory institutions, best able to capitalise on anticipated future improvements in values and information (Gaus 2018; Barrett 2020)?

Economics:

- Certain institutions, such as 'inclusive' governments, appear to be associated both with substantial increases in economic growth, across many generations, and with decreases in the probability of events (such as wars) that may be associated with catastrophic risk (e.g., Acemoglu et al. 2005; Huth and Allee 2002). How can we estimate the social benefits of various institution-building efforts in the long term?

- When faced with an important, irreversible decision for which relevant information will soon be obtained, a social planner may preserve 'option value' by delaying the decision until after the information has been acquired (cf. Bishop 1982; Dixit and Pindyck 1994). In delaying an important social decision intergenerationally, however, the planner may worry that future agents' values and preferences will differ from their own (INFORMAL: Hanson 2018). Facing this tradeoff, under what conditions should the planner – or principal – defer irreversible decisions to better-informed future agents (MacAskill, MS) (INFORMAL: Brauner and Grosse-Holz 2018)? Can long-term intergenerational mechanisms help to overcome this principal-agent problem? What might they look like (Bostrom 2006) (INFORMAL: Tomasik 2018)?

- How will future generations react to current altruistic efforts? While intergenerational altruistic efforts may be strategic substitutes (high effort today may induce low effort tomorrow), is it also possible that current actions will set a precedent that future generations are inclined to follow (Summers and Zeckhauser, 2008)?

Interdisciplinary:

- Do 'broad' approaches to improving the far future (such as promoting good institutions or global peace) tend to be more or less effective, in expectation, than 'narrow' approaches (such as working on reducing the risk of bioengineered pandemics) (INFORMAL: Beckstead 2013)?

- Is there a strong case for enfranchising future generations (Kavka and Warren 1983; Goodin 2007; Tännsjö 2007; Beckman 2009: Ch. 7)? If so, how should this be implemented? How effectively can contemporary individuals act as representatives for future generations? Might other democratic reforms, such as age-weighting, help to better align political outcomes with long-term priorities (Gonzalez-Ricoy and Gosseries 2017) (INFORMAL: MacAskill 2019)?

PHIL - POLITICAL PHILOSOPHY   ECON - POLITICAL ECONOMY, MECHANISM DESIGN

Existing academic literature:

- Acemoglu, Daron, Simon Johnson and James A. Robinson. 'Institutions as a Fundamental Cause of Long-Run Growth'. *Handbook of Economic Growth*, 1A (2005): 385–472.

- Barrett, Jacob.'Social Reform in a Complex World'. *Journal of Ethics and Social Philosophy* 17, no. 2 (2020): 103–132.

- Beckman, Ludvig. *The Frontiers of Democracy: The Right to Vote and Its Limits*. Basingstoke: Palgrave MacMillan, 2009.

- Bishop, Richard C. 'Option Value: An Exposition and Extension'. *Land Economics* 58, no. 1 (1982): 1–15.

- Bostrom, Nick. 'What is a Singleton?' *Linguistic and Philosophical Investigations* 5, no. 2 (2006): 48–54.

- Cotton-Barratt, Owen. 'Allocating Risk Mitigation across Time'. *Future of Humanity Institute*, Technical Report #2015-2.

- Dixit, Avinash and Robert S. Pindyck. *Investment Under Uncertainty*. Princeton: Princeton University Press, 1994.

- Estlund, David . *Utopophobia: On the Limits (If Any) of Political Philosophy*. Princeton, NJ: Princeton University Press, 2019.

- Gaus, Gerald. *The Tyranny of the Ideal: Justice in a Diverse Society*. Princeton, NJ: Princeton University Press, 2016.

- Gonzalez-Ricoy, Inigo and Gosseries, Axel, eds. *Institutions for Future Generations*. Oxford: Oxford University Press, 2017.

- Goodin, Robert E. 'Enfranchising All Affected Interests, And Its Alternatives'. *Philosophy and Public Affairs* 35, no. 1 (2007): 40–68.

- Huth, Paul K., and Todd L. Allee. *The Democratic Peace and Territorial Conflict in the Twentieth Century*. Vol. 82. Cambridge University Press, 2002.

- Kavka, Gregory and Warren, Virginia. 'Political Representation for Future Generations'. In *Environmental Philosophy: A Collection of Readings*, edited by Robert Elliott and Arran Gare. University Park, PA: the Pennsylvania State University Press, 1983.

- Kimball, Miles S. 'Making Sense of Two-Sided Altruism'. *Journal of Monetary Economics* 20, no. 2 (1987): 301–326.

- Lawford-Smith, Holly. 'Understanding Political Feasibility'. *Journal of Political Philosophy* 21, no. 3 (2013): 243–259.

- MacAskill, William. 'Human Extinction, Asymmetry, and Option Value'. Manuscript in preparation.

- Sen, Amartya. *The Idea of Justice*. Harvard, MA: Harvard University Press, 2009.

- Summers, Lawrence and Richard Zeckhauser. 'Policymaking for posterity'. *Journal of Risk and Uncertainty* 37 (2008): 115–140.

- Tännsjö, Torbjörn. 'Future People, the All Affected Principle, and the Limits of the Aggregation Model of Democracy'. In *Homage a Wlodek: Philosophical Papers Dedicated to Wlodek Rabinowicz*, edited by Toni Rønnow-Rasmussen, Björn Petersson, Jonas Josefsson, and Dan Egonsson. Lund, 2007.

Existing informal discussion:

- Nick Beckstead, How to compare broad and targeted attempts to shape the far future, 2013.

- Jan Brauner and Friederike Grosse-Holz, The expected value of extinction risk reduction is positive, 2018.

- Robin Hanson, On value drift, 2018.

- William MacAskill, Our descendants will probably see us as moral monsters. What should we do about that?, 2018.

- William MacAskill, Age-Weighted Voting, 2019.

- Toby Ord, The timing of labour aimed at reducing existential risk, 2014.

- Brian Tomasik, Will Future Civilization Eventually Achieve Goal Preservation?, 2018.

1.6. Economic growth, population growth and inequality

In the past two hundred years, GDP per capita has increased by more than a factor of 20 in Western countries, leading to increases in consumption, standards of living and better health outcomes (Jones 2016b). From a longtermist perspective, increasing economic growth could therefore be extremely beneficial as it promises to improve the entire course of the future. However, there are several reasons to think that economic growth might fail to deliver on this promise. Research to improve understanding of the benefits, costs, risks and limits of economic growth may therefore be important in setting global priorities.

Economics:

- What is the relationship between long-term economic growth and human welfare? How well does GDP per capita serve as a proxy for human welfare (Stevenson and Wolfers 2008; Klenow and Jones 2016; Easterlin 2016)? Are increases to human welfare from accelerating economic growth fundamentally limited, for example by

an upper bound to output per capita or by an upper bound to welfare as a function of consumption (INFORMAL: Trammell 2019)?

- What is the long-term relationship between economic growth and desirable outcomes other than human welfare, for example the welfare of non-human animals (Frank 2008; Olsson and Alexandrie 2019)?

- Endogenous growth models usually assume a growing or constant population. Is (global) population growth a requisite of long run economic growth? What does that imply for optimal population dynamics in the long run? What are the consequences of a declining population for long run economic growth and social welfare (Jones 2020)?

- Is the global economy likely to transition into a new growth mode, with a significantly higher long-term growth rate (Hanson 2000, Roodman 2020), or a significantly lower growth rate (Gordon 2014)? Are there reasons to think that rapid improvements in artificial intelligence will result in a significantly higher growth rate (Hanson 2001; Nordhaus 2015; Aghion et al. 2017; Trammell 2020)?

- Some economists have argued that the welfare gains from accelerating growth far outstrip the benefits from redistribution (Cowen 2018). For example, under plausible assumptions, economic growth can cause *welfare* inequality to decline even if absolute *wealth* inequality persists. However, economic inequality may reduce growth, for example as a result of under-utilisation of human capital or sub-optimal distribution of political power (Stiglitz 2013), and may be self-reinforcing in the long run. To what extent should longtermists aim to reduce economic inequality?

- What are the most cost-effective ways to foster long-run global economic growth (INFORMAL: Halstead and Hillebrandt 2020)? Does this involve lobbying governments for pro-growth policies, investing in productive capital or R&D? How do these compare to other efforts to increase global welfare?

- Of the comprehensive macroeconomic indices already available to us, which serve best as proxies for long-term expected global welfare (Cockburn et al. 2014), including but not limited to considerations of existential risks? What would be the broad policy implications of targeting such indices instead of GDP per capita?

- Are there any promising proxies for long-term wellbeing *not* already tracked as macroeconomic indices (INFORMAL: Shulman 2013; Bostrom 2014)? If so, how could these proxies be formalised and measured, and what would be the broad policy implications of targeting them instead of GDP per capita?

Interdisciplinary:

- Wealth is distributed across individuals who vary across many dimensions, including their moral values and patience. How is the distribution of wealth across moral values likely to evolve in the future (INFORMAL: Trammell 2020)? Are longtermist

agents likely to control a relatively larger fraction of future resources by virtue of their patience (INFORMAL: Christiano 2013)?

ECON - GROWTH, EVOLUTIONARY GAME THEORY, DISCOUNTING

- What population size is optimal from a longtermist perspective (Greaves forthcoming)?

PHIL - POPULATION ETHICS   ECON - POPULATION, DEMOGRAPHY

Existing academic literature:

- Aghion, Philippe, Benjamin F. Jones and Charles I. Jones. 'Artificial intelligence and Economic Growth'. No. w23928. National Bureau of Economic Research, 2017.

- Cockburn, John, Jean-Yves Duclos and Agnès Zabsonré. 'Is global social welfare increasing? A critical-level enquiry.' *Journal of Public Economics* 118 (2014): 151–162.

- Cowen, Tyler. 'Caring About the Distant Future: Why It Matters and What It Means'. *The University Of Chicago Law Review* 74, no. 5 (2007): 5–40.

- ———. *Stubborn Attachments*. San Francisco: Stripe Press, 2018.

- Easterlin, Richard A. 'Paradox lost?' USC-INET Research Paper 16-02 (2016).

- Frank, Joshua. 'Is there an "animal welfare Kuznets curve"?.' *Ecological Economics* 66.2-3 (2008): 478–491.

- Friedman, Benjamin M. *The Moral Consequences of Economic Growth*. New York: Vintage Books, 2006.

- Gordon, Robert J. 'The Demise of US Economic Growth: Restatement, Rebuttal, and Reflections'. No. w19895. National Bureau of Economic Research, 2014.

- Greaves, Hilary. 'Optimum Population Size'. In *Oxford Handbook of Population Ethics*, edited by Gustaf Arrhenius, Krister Bykvist and Tim Campbell. Oxford: Oxford University Press, forthcoming.

- Hanson, Robin. 'Long-term growth as a sequence of exponential modes'. Working manuscript, 2000.

- ———. 'Economic growth given machine intelligence'. Technical Report, University of California, Berkeley, 2001.

- Klenow, Peter J. and Chad I. Jones. 'Beyond GDP? Welfare across Countries and Time'. *American Economic Review* 106, no. 9 (2016): 2,426–2,457

- Jones, Charles I. 'Life and Growth'. *Journal of Political Economy* 124, no. 2 (2016a): 539–578.

- ———. 'The Facts of Economic Growth'. *Handbook of Macroeconomics*, 2A (2016b): 3–69.

- ———. 'The End of Economic Growth? Unintended Consequences of a Declining Population'. Working paper, 2020.

- Nordhaus, William J. 'Are We Approaching an Economic Singularity? Information Technology and the Future of Economic Growth'. Cowles Foundation Discussion Paper 2021, 2015.

- Olsson, Caroline and Gustav Alexandrie. '[A Beef with Growth? An Empirical Analysis of Income and Animal Farming](#)'. Bachelor Thesis, Stockholm University 2019.

- Roodman, David. '[On the probability distribution of long-term changes in the growth rate of the global economy: An outside view](#)'. Working paper, 2020.

- Stevenson, Betsey, and Justin Wolfers. '[Economic growth and subjective well-being: Reassessing the Easterlin paradox](#)'. *Brookings Papers on Economic Activity* 39, no. 1 (2008): 1–102.

- Stiglitz, Joseph E. *The Price of Inequality: How Today's Divided Society Endangers our Future.* London: Penguin Books, 2013.

- Trammell, Philip and Anton Korinek. '[Economic Growth Under Transformative AI](#)'. Working paper, 2020.

Existing informal discussion:

- Nick Beckstead, [How much can a long-run perspective help with strategic cause selection?](#), 2014.

- Nick Bostrom, [Crucial considerations and wise philanthropy](#), 2014.

- Kyle Bogosian, [Quantifying the Impact of Economic Growth on Meat Consumption](#), 2015.

- Paul Christiano, [Why might the future be good?](#), 2013.

- John Halstead and Hauke Hillebrandt, [Growth and the case against randomista development](#), 2020.

- Carl Shulman, [What proxies to use for flow-through effects?](#), 2013.

- Philip Trammell, [Welfare Implications of Accelerating Long-Run Consumption Growth](#), 2019.

- Philip Trammell, [The Enemy of the Good](#), 2020.

- Brian Tomasik, [Differential Intellectual Progress as a Positive-Sum Project](#) ('Economic growth' subsection), 2013.

- Eliezer Yudkowsky, [Do Earths with slower economic growth have a better chance at FAI?](#), 2013.

1.7. Moral uncertainty for longtermists

Estimates of the value of an intervention are sensitive not only to uncertainty about the intervention's consequences, but also to uncertainty about the normative criteria by which to evaluate its consequences. This 'moral uncertainty' may prove particularly important from a longtermist perspective, as we may often have to choose between interventions whose short-term consequences are of similar value across all plausible moral theories, but whose long-term consequences differ substantially across plausible moral theories.

Philosophy:

- Are there instrumental goals on which competing axiologies converge? Given axiological uncertainty, can we make any claims about what sort of future we should try to aim for (Greaves and Ord 2017; MacAskill 2019) (INFORMAL: Pummer 2015; Leech 2018)?

PHIL - MORAL UNCERTAINTY

- How likely is it that civilisation will converge on the correct moral theory given enough time? What implications does this have for cause prioritisation in the nearer term (Gustafsson and Peterson 2012; Vallinder and Olsson 2013; MacAskill MS)?

PHIL - MORAL UNCERTAINTY

- How likely is it that the correct moral theory is a theory radically different from any yet proposed? If likely, how likely is it that civilisation will discover it, and converge on it, given enough time? While it remains unknown, how can we properly hedge against the associated moral risk (MacAskill MS)?

PHIL - MORAL UNCERTAINTY

- Under moral uncertainty, do some axiological views with very high stakes swamp the expected value calculation? If so, which views are they? What is the best way to deal with this 'fanaticism' issue (Ross 2006; MacAskill and Ord 2018; Cotton-Barratt and Greaves 2019) (INFORMAL: MacAskill 2018)?

PHIL - MORAL UNCERTAINTY, DECISION THEORY

Interdisciplinary:

- Given the informed disagreements that persist regarding the social discounting rate and other normative issues of relevance for longtermism, non-dogmatic agents should arguably be uncertain about these issues. What social discount rate and other normative assumptions should be adopted in light of this uncertainty (Millner 2020; Jaakkola and Millner 2020)?

PHIL - MORAL UNCERTAINTY, DECISION THEORY   ECON - DECISION THEORY, WELFARE ECONOMICS

Existing academic literature:

- Greaves, Hilary, and Toby Ord. 'Moral Uncertainty About Population Axiology'. *Journal of Ethics & Social Philosophy* 12, no. 2 (2017): 135–167.

- Greaves, Hilary and Owen Cotton-Barratt. 'A Bargaining-theoretic approach to moral uncertainty'. Working paper, 2019.

- Gustafsson, Johan E. and Martin Peterson. 'A computer simulation of the argument from disagreement.' *Synthese* 184.3 (2012): 387–405.

- Jaakkola, Niko and Antony Millner. 'Non-dogmatic climate policy'. Working paper, 2020.

- MacAskill, William. 'Practical Ethics Given Moral Uncertainty'. *Utilitas* 31, no. 3 (2019): 231–245.

- ———. 'Should we expect moral convergence?'. Manuscript in preparation.

- MacAskill, William and Toby Ord. '[Why Maximize Expected Choice-Worthiness?](#)' *Noûs*, 14 July 2018.

- Millner, Antony. '[Nondogmatic Social Discounting](#)'. *American Economic Review* 110.3 (2020): 760–775.

- Ross, Jacob. '[Rejecting Ethical Deflationism](#)'. *Ethics* 116, no. 4 (2006): 742–768.

- Vallinder, Aron and Erik J. Olsson. '[Do computer simulations support the Argument from Disagreement?](#)' *Synthese* 190.8 (2013): 1,437–1,454.

Existing informal discussion:

- John Halstead, [Moral uncertainty and climate change](#), 2017.

- Holden Karnofsky, [Update on Cause Prioritization at Open Philanthropy](#), 2018.

- Gavin Leech, [Existential risk as common cause](#), 2018.

- William MacAskill, [Our descendants will probably see us as moral monsters. What should we do about that?](#), 2018.

- Theron Pummer, [Moral Agreement on Saving the World](#), 2015.

1.8. Forecasting the long-term future

It is sometimes argued that we cannot *predictably* influence the long-term future and therefore the *expected* value of our actions is primarily determined by short-term considerations, even if one accepts that the future holds enormous potential value. In light of this important objection to longtermism, it is valuable to learn more about our current ability to forecast the long-term effects of our actions, how it could be improved, and which type of long-term forecasts are particularly relevant from a longtermist perspective. Research on long-term forecasting could also help us to compare the expected value of different strategies for improving the long-term future and to determine the extent to which longtermists should invest in better forecasts.

Philosophy:

- Forecasting the long-term effects of our actions often requires us to make difficult comparisons between complex and messy bodies of competing evidence, a situation Greaves (2016) calls "complex cluelessness". We must also reckon with our own incomplete awareness, that is, the likelihood that the long-run future will be shaped by events we've never considered and perhaps can't fully imagine. What is the appropriate response to this sort of epistemic situation? For instance, does rationality require us to adopt precise subjective probabilities concerning the very-long-run effects of our actions, imprecise probabilities (and if so, how imprecise?), or some other sort of doxastic state entirely?

- In recent centuries, revolutionary ideas like evolutionary theory and computing have vastly reshaped both our understanding of the physical world and the actual workings of the social world. It is natural to suppose that other similarly transformative ideas will be discovered in coming centuries. Does our ignorance of

future transformative ideas constitute a major obstacle to predicting and influencing the far future, as Deutsch (2011) suggests? If so, how should we respond? To what extent, if at all, does this undercut the case for longtermism?

- Which current gaps in our knowledge regarding the very long term are particularly action relevant? In which scientific field or other domain could these gaps be closed by an increase in (reliable) long-term forecasts?

Economics:

- In at least some modelling frameworks, short-term predictability may be more important than long-run forecasting (Millner and Heyen 2019). In which settings do the conditions of these models apply, and how widely applicable is this conclusion?

- Is it possible to incentivise accuracy in long-term forecasts, given that we may not live to see the realisation of the event we wish to forecast? Can we make use of incentives to report 'subjective truths' (Prelec 2004) for this purpose?

- Forecasts of standard econometric time series models will often collapse into simple predictions when applied to the far future (Granger and Jeon 2007). Are these predictions still informative for longtermist decision-making? Can econometric tools be adjusted to make more substantive long-run forecasts?

- How accurate have forecasts about long-term developments been in the past? Are efforts to 'predict history' (Risi et al. 2019) informative about our ability to predict the long-run future?

Interdisciplinary:

- It is plausible that, in general, our ability to predictably influence the state of the world decreases with time (e.g., because of the continual possibility of exogenous events that wipe out the effects of our actions), but the importance of influencing the state of the world increases with time (e.g., because our civilisation grows and its potential for value increases). Which of these effects is quantitatively more important in the long run? How much does the ongoing possibility of catastrophes and other major exogenous events in the far future reduce the expected value of longtermist interventions (Tarsney 2019; Ord 2020, Appendix E)?

- The performance of human forecasting in the short- to medium-term (from years to decades) has only recently been systematically tracked (Tetlock and Gardner 2016; Tetlock 2017). Existing research has drawn heavily on psychology and behavioural economics to determine common biases in human approaches to thinking about the future and how to overcome them. Does the existing data already allow for direct evaluation of human forecasting capabilities on longer time scales? Can we reliably extrapolate findings from factors that influence short-term forecasting success to

forecasts of the very long term? How well do alternative approaches, such as prediction markets (Arrow et al. 2008; Hanson 2013) or the Delphi method (Helmer 1967), perform in comparison?

ECON - HEURISTICS AND BIASES, INTUITIVE FORECASTING   PHIL - JUDGMENT AGGREGATION

- What long-run forecasts are implied by major theories in the social sciences, in particular by economic theory (e.g., models of growth or political change)? Should long term implications of major theories be treated as empirical predictions? If so, are they supported by evidence, and are they compatible with other predictions about the long term future?

ECON - GROWTH, POLITICAL ECONOMY   PHIL - PHILOSOPHY OF SCIENCE

- To what extent is the short-term predictive track record of theories, of the type promoted in recent forecasting competitions (Camerer et al. 2016; DellaVigna and Pope 2018; Dreber et al. 2015; Erev et al. 2010) informative of long-term predictive success?

ECON - APPLIED ECONOMETRICS, EXPERIMENTAL ECONOMICS, FIELD EXPERIMENTS   PHIL - PHILOSOPHY OF SCIENCE

Existing academic literature:

- Camerer, Colin F., Anna Dreber, Eskil Forsell, Teck-Hua Ho, Jürgen Huber, Magnus Johannesson, Michael Kirchler, Johan Almenberg, Adam Altmejd, Taizan Chan, Emma Heikensten, Felix Holzmeister, Taisuke Imai, Siri Isaksson, Gideon Nave, Thomas Pfeiffer, Michael Razen and Hang Wu 'Evaluating replicability of laboratory experiments in economics'. *Science* 351.6280 (2016): 1,433–1,436.

- Christensen, Peter, Kenneth Gillingham and William Nordhaus. 'Uncertainty in forecasts of long-run economic growth.' *Proceedings of the National Academy of Sciences* 115.21 (2018): 5,409–5,414.

- DellaVigna, Stefano and Devin Pope. 'Predicting experimental results: who knows what?' *Journal of Political Economy* 126.6 (2018): 2,410–2,456.

- Deutsch, David. *The beginning of infinity: Explanations that transform the world.* Penguin UK, 2011.

- Dreber, Anna, Thomas Pfeiffer, Johan Almenberg, Siri Isaksson, Brad Wilson, Yiling Chen, Brian A. Nosek and Magnus Johannesson. 'Using prediction markets to estimate the reproducibility of scientific research'. *Proceedings of the National Academy of Sciences* 112.50 (2015): 15,343–15,347.

- Erev, Ido, Eyal Ert, Alvin E. Roth, Ernan Haruvy, Stefan M. Herzog, Robin Hau, Ralph Hertwig, Terrence Stewart, Robert West and Christian Lebiere. 'A choice prediction competition: Choices from experience and from description'. *Journal of Behavioral Decision Making* 23.1 (2010): 15–47.

- Granger, Clive WJ and Yongil Jeon. 'Long-term forecasting and evaluation'. *International Journal of Forecasting* 23.4 (2007): 539–551.

- Greaves, Hilary. 'Cluelessness'. *Proceedings of the Aristotelian Society* 116 (2016): 311–339.

- Kott, Alexander and Philip Perconti. 'Long-term forecasts of military technologies for a 20–30 year horizon: An empirical assessment of accuracy'. *Technological Forecasting and Social Change* 137 (2018): 272–279.

- Lempert, Robert J. Popper, Steven W. and Bankes, Steven C. 'Shaping the next one hundred years: new methods for quantitative, long-term policy analysis.' RAND Corporation, 2003.

- Müller, Ulrich K. and Mark W. Watson. 'Measuring uncertainty about long-run predictions'. *Review of Economic Studies* 83.4 (2016): 1,711–1,740.

- National Research Council. 'The Accuracy of Past Projections'. Chapter 2 in: *Beyond six billion: Forecasting the world's population*. National Academies Press, 2000.

- Ord, Toby. *The Precipice: Existential Risk and the Future of Humanity*. London: Bloomsbury, 2020.

- Risi, Joseph, Amit Sharma, Rohan Shah, Matthew Connelly and Duncan J. Watts. 'Predicting history.' *Nature Human Behaviour* 3.9 (2019): 906–912.

- Tarsney, Christian. 'The Epistemic Challenge to Longtermism.' Working paper, 2019.

- Tetlock, Philip E. *Expert Political Judgment: How Good Is It? How Can We Know?-New Edition*. Princeton University Press, 2017.

- Tetlock, Philip E. and Dan Gardner. *Superforecasting: The Art and Science of Prediction*. Random House, 2016.

Existing informal discussion:

- Luke Muehlhauser, How Feasible Is Long-range Forecasting, 2019.

2. General issues in global prioritisation

There is a large set of theoretical and empirical questions that arise for agents engaged in global prioritisation that are not specific to the longtermist point of view. These include questions about evidence aggregation, optimal timing and coordination for altruistic agents. The following areas of research strike us as particularly interesting and important.

2.1. Decision-theoretic issues

The framework of expected utility theory sometimes produces deeply counterintuitive conclusions, especially in situations involving extremely low-probability, high-magnitude payoffs. When faced with the possibility of infinite payoffs, the expected utility framework breaks down altogether. These and other decision-theoretic problems are of particular interest to individuals or organisations trying to do good, whose concerns may extend beyond the relatively local scope for which standard decision theory has been developed, and warrant the development of nonstandard decision-theoretic solutions.

Philosophy:

- Faced with the task of comparing actions in terms of expected value, it often seems that the agent is 'clueless': that is, that the available empirical and theoretical evidence simply supplies too thin a basis for guiding decisions in any principled way (Lenman 2000; Greaves 2016; Mogensen 2020) (INFORMAL: Tomasik 2013; Askell 2018). How is this situation best modelled, and what is the rational way of making decisions when in this predicament? Does cluelessness systematically favour some types of action over others?

PHIL - DECISION THEORY

- Should an actor have a prior belief over the distribution of her possible impact (INFORMAL: Karnofsky 2011) such that it is astronomically unlikely that she could have the sort of positive impact that it seems one can have by reducing existential risk if total utilitarianism is correct? What bearing does this have on the expected value of activities aiming to improve the long-run future?

PHIL - DECISION THEORY, EPISTEMOLOGY

- Often it seems that subtle differences in epistemology would lead one to quite different conclusions concerning which interventions have the highest expected impartial value. These include differences in responses to paucity of hard evidence, in level of trust in abstract arguments leading to counterintuitive conclusions, in responses to interpersonal disagreement, and in the relative weight placed on different types of evidence. To what extent should this lack of robustness move us away from simply maximising expected value with respect to whatever credences we happen (now) to have? Is there a plausible alternative approach (INFORMAL: Karnofsky 2016)?

PHIL - DECISION THEORY, EPISTEMOLOGY

Interdisciplinary:

- Should we favour interventions that have more evidential support, all else being equal? If so, this may conflict with expected utility maximisation if one would prefer an intervention with much stronger evidence but a (possibly infinitesimally) lower expected value. On the other hand, it also seems reasonable to place some value on the uncertainty of an intervention. What is the correct response to this tradeoff (Askell 2019) (INFORMAL: Hurford 2013)?

PHIL - EPISTEMOLOGY, DECISION THEORY   ECON - VALUE OF INFORMATION

- If most interventions are indeed fairly ineffective, is it the case that interventions that are supported only by speculative evidence will generally have lower *expected* value than that of interventions supported by more solid evidence?

PHIL - DECISION THEORY, EPISTEMOLOGY   ECON - VALUE OF INFORMATION, BAYESIAN DECISION THEORY

- To what extent should we be risk averse in our approach to doing good, and what are the implications of reasonable risk aversion for global prioritisation? (Quiggin 1982; Buchak 2013; Greaves et al. MS)

PHIL - DECISION THEORY   ECON - DECISION THEORY

- What are the implications of ambiguity aversion (whether rational or not) for the project of doing good (Rowe and Voorhoeve 2018)?

PHIL - DECISION THEORY   ECON - DECISION THEORY

Existing academic literature:

- Askell, Amanda. 'Evidence Neutrality and the Moral Value of Information'. In *Effective Altruism: Philosophical Issues*, edited by Hilary Greaves and Theron Pummer. Oxford: Oxford University Press, 2019.

- Bostrom, Nick. 'Pascal's Mugging'. *Analysis* 69, no. 3 (2009): 443–445.

- Buchak, Lara. *Risk and Rationality*. Oxford: Oxford University Press, 2013.

- Greaves, Hilary. 'Cluelessness'. *Proceedings of the Aristotelian Society* 116 (2016): 311–339.

- Greaves, Hilary, Andreas L. Mogensen and William MacAskill. 'Longtermism for Risk Averse Altruists'. Manuscript in preparation.

- Jansen, Christoph, Georg Schollmeyer and Thomas Augustin. 'Concepts for Decision Making under Severe Uncertainty with Partial Ordinal and Partial Cardinal Preferences'. *International Journal of Approximate Reasoning* 98 (2018): 112–131.

- Lenman, James. 'Consequentialism and Cluelessness'. *Philosophy & Public Affairs* 29, no. 4 (2000): 342–370.

- Mogensen, Andreas L. 'Maximal cluelessness'. *The Philosophical Quarterly* (2020).

- Quiggin, John. 'A Theory of Anticipated Utility'. *Journal of Economic Behavior & Organization* 3, no. 4 (1982): 323–343.

- Tarsney, Christian. 'Exceeding Expectations: Stochastic Dominance as a General Decision Theory'. Working paper, 2019.

Existing informal discussion:

- Amanda Askell, [Tackling the ethics of infinity, being clueless about the effects of our actions, and having moral empathy for intellectual adversaries](), 2018.

- Amanda Askell, [Moral Value of Information](), 2017.

- Tobias Baumann, [Uncertainty smooths out differences in impact](), 2017.

- Peter Hurford, [Why I'm Skeptical About Unproven Causes (And You Should Be Too)](), 2013.

- Holden Karnofsky, [Why we can't take expected value estimates literally (even when they're unbiased)](), 2011.

- Holden Karnofsky, [Worldview diversification](), 2016.

- Brian Tomasik, [Charity cost-effectiveness in an uncertain world](), 2013.

2.2. Epistemological issues

Thinking about global prioritisation, particularly (although not only) within the longtermist paradigm, tends to rely on heavily philosophical considerations and to reach some surprising and counterintuitive conclusions. We must therefore assess the extent to which this unusual circumstance should undermine our confidence in the conclusions in question.

Philosophy:

- To what extent should an actor place weight on her own personal assessment of the first-order evidence, rather than deferring to the views of peers or experts (Elga 2007; Christensen 2007; Christensen 2009; Feldman and Warfield 2010; Wilson 2010; Christensen and Lackey 2013) (INFORMAL: Beckstead 2013; Lewis 2017)?

PHIL - MORAL UNCERTAINTY ECON - VALUE OF INFORMATION, BAYESIAN UPDATING

- How much weight should we place on philosophical arguments? Is there a sound 'pessimistic induction' against placing much weight on them, assuming that most philosophical arguments in the past have been mistaken?

PHIL - EPISTEMOLOGY, METAPHILOSOPHY

- Should one have the same levels of epistemic modesty about unusual moral views as one should about unusual empirical views?

PHIL - EPISTEMOLOGY, MORAL UNCERTAINTY

Existing academic literature:

- Christensen, David. '[Epistemology of Disagreement: The Good News]()'. *Philosophical Review* 116, no. 2 (2007): 187–217.

- ———. '[Disagreement as Evidence: The Epistemology of Controversy]()'. *Philosophy Compass* 4, no. 5 (2009): 756–767.

- Christensen, David and Jennifer Lackey, eds. *The Epistemology of Disagreement: New Essays*. Oxford: Oxford University Press, 2013.

- Elga, Adam. '[Reflection and Disagreement](#)'. *Noûs* 41, no. 3 (2007): 478–502.
- Feldman, Richard, and Ted A. Warfield, eds. *Disagreement*. Oxford: Oxford University Press, 2010.
- Wilson, Alastair. '[Disagreement, Equal Weight, and Commutativity](#)'. *Philosophical Studies* 149 (2010): 321–326.

Existing informal discussion:

- Amanda Askell, [The Moral Value of Information](#), 2017.
- Nick Beckstead, [Common sense as a prior](#), 2013.
- Holden Karnofsky, [Maximising cost-effectiveness via critical enquiry](#), 2011.
- Holden Karnofsky, [Sequence thinking vs. cluster thinking](#), 2014.
- Holden Karnofsky, [Modelling Extreme Model Uncertainty](#), 2014.
- Greg Lewis, [In defence of epistemic modesty](#), 2017.
- Jonah Sinick, [Many Weak Arguments vs. One Relatively Strong Argument](#), 2013.
- Benjamin Todd, [Is it fair to say that most social programmes don't work?](#), 2017.

2.3. Optimal timing and discounting

An agent aiming to do good faces two central timing questions.

First: When should she put her resources to philanthropic use? Should she make donations as she earns money, invest to donate at a later date, or borrow to donate more now? Should she take a high impact job today, or invest in human and career capital? Do the conclusions vary for different types of agents (e.g., individuals, foundations, governments) depending on the scale of resources available to them and the constraints that they face (also see [Section 2.8](#))?

Second: How should she compare payoffs which will accrue at different time periods?

Philosophy:

- Some (e.g., Parfit 2011) have argued that the present is an unusual time with respect to how quickly we ought to discount future donations. Is this correct?

PHIL - ETHICS OF DISCOUNTING

Economics:

- Does the analysis of optimal timing depend on whether we are considering monetary investments or investments in human capital? What relevant considerations apply in one case but not the other? For example, it is much more difficult to 'borrow' human capital than it is to borrow for a monetary investment.

ECON - DYNAMIC OPTIMISATION, HUMAN CAPITAL

- Does uncertainty over the long run effects of interventions in a specific cause area provide a reason to favour investments that can be used more flexibly across interventions in the future? For instance, should altruistic agents learn broadly useful skills, instead of specialising in a field with short run payoffs but uncertain long term prospects (Cotton-Barratt 2015) (INFORMAL: Ord 2014)?

ECON - DISCOUNTING, HUMAN CAPITAL

- It is often observed that smaller donors tend to give as they earn, while larger donors and institutions save and give later. Why is this? Is this optimal? If not, can policy improve the allocation of donations over time?

ECON - DISCOUNTING, PUBLIC ECONOMICS

- Private investors typically discount monetary returns not only for temporally neutral reasons, such as the prospect of higher personal consumption in the future, but also for reasons of pure time preference. As a result, market interest rates should be expected to exceed the rate at which the marginal utility of consumption is declining. Does this imply that, under ordinary circumstances, temporally neutral altruists should save rather than give?

ECON - DISCOUNTING, INTERGENERATIONAL EQUITY, FINANCIAL ECONOMICS

- Policymakers typically discount dollar-valued social costs and benefits not only for temporally neutral reasons, such as the prospect of higher average consumption in the future, but also as a reflection of citizens' pure time preferences and short-term political incentives. How would policy recommendations change on evaluating social costs and benefits from a temporally neutral perspective (as explored in the optimal taxation context, for example, by Barrage (2018))? How might a patient agent provide incentives for an impatient government to implement policy consistent with placing a higher valuation on the future? How might an unusually patient government provide incentives for future (possibly impatient) governments to continue to make future-oriented investments?

ECON - DISCOUNTING, POLITICAL ECONOMY, INTERGENERATIONAL EQUITY, INTERGENERATIONAL GOVERNANCE, MECHANISM DESIGN

- How should a patient philanthropist spend her money to account for current and emerging existential risks?

ECON - DISCOUNTING, CATASTROPHIC RISK, INTERGENERATIONAL EQUITY

- What are the key considerations that determine the optimal timing of philanthropic giving or social spending? Positive returns on investment and increasing information about the effectiveness of interventions are important reasons to wait before committing resources to philanthropic use. On the other hand, economic growth, and with it declining opportunities for cost-effective philanthropy, are important reasons to put resources to use earlier. How are optimal timing decisions best modelled (Weitzman 1998; Gollier 2004; Fleurbaey and Zuber 2015; Trammell MS; MacAskill 2019) (INFORMAL: Christiano 2013a; Christiano 2013b; Wise 2013; Cotton-Barratt and Todd 2015; Todd 2017)? Does search theory (e.g., Mortensen 1986) or the analysis of optimal stopping problems (e.g., Rust 1987) provide a useful framework for such decisions?

ECON - DISCOUNTING, VALUE OF INFORMATION, FORECASTING

Interdisciplinary:

- Some consumption today consists of activity that is likely of negative value, such as inhumane animal agriculture. Other consumption, such as pain relief, may have unusually large positive value compared to what market prices suggest. How severely do such considerations render Ramsey-discounted consumption an imperfect proxy for moral value? How should we expect the weight of such considerations to change in the future?

Existing academic literature:

- Barrage, Lint. 'Be Careful What You Calibrate for: Social Discounting in General Equilibrium'. *Journal of Public Economics* 160 (2018): 33–49.

- Cotton-Barratt, Owen. 'Allocating Risk Mitigation across Time'. *Future of Humanity Institute*, Technical Report #2015-2.

- Emerson, Jed, Jay Wachowicz and Suzi Chun. 'Social Return on Investment: Exploring Aspects of Value Creation in the Nonprofit Sector'. *Social Purpose Enterprises and Venture Philanthropy in the New Millennium* 2 (2000): 132–173.

- Fleurbaey, Marc and Stéphane Zuber. 'Discounting, Risk and Inequality: A General Approach'. *Journal of Public Economics* 128 (2015): 34–49.

- Frumkin, Peter. *Strategic Giving: The Art and Science of Philanthropy*. Chicago: University of Chicago Press, 2016.

- Gollier, Christian. 'Maximizing the Expected Net Future Value as an Alternative Strategy to Gamma Discounting'. *Finance Research Letters* 1, no. 2 (2004): 85–89.

- Greaves, Hilary. 'Discounting for Public Policy: A Survey'. *Economics & Philosophy* 33, no. 3 (2017): 391–439.

- Irvin, Renée A. 'Endowments: Stable Largesse or Distortion of the Polity?' *Public Administration Review* 67, no. 3 (2007): 445–457.

- Jansen, Paul and David Katz. 'For Nonprofits, Time Is Money'. *The McKinsey Quarterly* 1 (2002): 124–133.

- Klausner, Michael D. 'When Time Isn't Money: Foundation Payouts and the Time Value of Money'. *SSRN Electronic Journal*, 2003.

- Landesman, Cliff. 'When to Terminate a Charitable Trust?' *Analysis* 55, no. 1 (1995): 12–13.

- MacAskill, William. 'When Should an Effective Altruist Donate?' Working paper, 2019.

- Méjean, Aurélie, Antonin Pottier, Stéphane Zuber and Marc Fleurbaey. 'Intergenerational Equity under Catastrophic Climate Change'. Working paper, 2017.

- Moller, Dan. 'Should We Let People Starve - for Now?' *Analysis* 66, no. 3 (2006): 240–247.

- Mortensen, Dale. 'Job Search and Labor Market Analysis'. *Handbook of Labor Economics* 2 (1986): 849–919.

- Nordhaus, William D. 'A Review of the *Stern Review on the Economics of Climate Change*'. *Journal of Economic Literature* 45, no. 3 (2007): 686–702.

- Parfit, Derek. *On What Matters*. Oxford: Oxford University Press, 2011.

- Rust, John. 'Optimal replacement of GMC bus engines: An empirical model of Harold Zurcher'. *Econometrica* 55, no. 5 (1987): 999–1,033.

- Stern, Nicholas. *The Economics of Climate Change*. Cambridge, UK: Cambridge University Press, 2006.

- Tarsney, Christian. 'Does a Discount Rate Measure the Costs of Climate Change?' *Economics & Philosophy* 33, no. 3 (2017): 337–365.

- Trammell, Philip. 'Patience and Philanthropy'. Manuscript in preparation.

- Weitzman, Martin L. 'Why the Far Distant Future Should Be Discounted at Its Lowest Possible Rate'. *Journal of Environmental Economics and Management* 36, no. 3 (1998): 201–208.

Existing informal discussion:

- Paul Christiano, Giving now vs. later, 2013a.

- Paul Christiano, The best reason to give later, 2013b.

- Owen Cotton-Barratt and Benjamin Todd, Give now or later? What to do when the order of your actions matters, 2015.

- Robin Hanson, Parable of the Multiplier Hole, 2010.

- Toby Ord, The timing of labour aimed at reducing existential risk, 2014.

- Benjamin Todd, Should you wait to make a difference?, 2017.

- Julia Wise, Giving now vs. later: a summary, 2013.

2.4. Diversification and hedging

What reasons are there, either for an individual philanthropist or for the global community of philanthropic actors, to diversify across causes/interventions, rather than simply identifying the intervention with the highest expected cost-effectiveness and supporting exclusively that intervention? Likewise, what reasons are there for philanthropic investors to diversify or hedge, instead of simply choosing the investments with highest expected return?

Philosophy:

- To what extent should a large foundation diversify across different 'worldviews' (INFORMAL: Karnofsky 2016)? To what extent does moral uncertainty provide support for such diversification?

Economics:

- Within the cause areas judged to be of exceptionally high priority, how quickly do we expect the returns to philanthropic resources to diminish (INFORMAL: Shulman 2014)? Given uncertainty about this rate (INFORMAL: Christiano 2013), should an investing philanthropist therefore sign 'charitable discount rate swaps', paying a sum if her discount rate is higher than expected (e.g., if some vaccine is developed more quickly than expected), in exchange for payment if it is lower? What other financial instruments might be used to hedge philanthropic risks and how might these be implemented?

- While investing for future giving, philanthropists may be able to maximise their impact by hedging their investments appropriately (Tran 2019). For example, an organisation with an environmental mission might hedge by investing in oil companies: in the case that fossil fuels become unexpectedly profitable (e.g., because of discoveries of large new oil reserves), the organisation will then have more resources available to make low carbon investments. How important is 'mission hedging' and how might it best be implemented?

- Mission hedging may also be possible for philanthropists with more open-ended missions. For example, one might think that, under most choices of cause area, philanthropic resources will go further when the market is doing poorly. In that case, market beta – an asset's excess return per unit of market excess return, as given by the Capital Asset Pricing Model (Sharpe 1964; Lintner 1965) – is serving as a proxy for 'philanthropic beta' – the association between the asset's value and the ease with which resources can be put to doing good. How well does market beta serve as a good proxy for philanthropic beta in the face of cause uncertainty? Might other market indices serve as better proxies?

- What are the potential reasons for diversifying investments across philanthropic causes? Possible justifications include: diminishing marginal returns of resources to impartial value within a given cause area or intervention, the information value of executing interventions and moral uncertainty. Which, if any, validly apply to individuals? To large foundations? To the worldwide community of altruistic actors as a whole (Snowden 2019; INFORMAL: Shulman 2012; Tomasik 2013)?

- How, if at all, do the considerations for or against diversifying across philanthropic causes differ when we consider how to allocate human capital resources rather than financial resources?

Existing academic literature:

- Elton, Edwin, Martin Gruber, Stephen Brown and William Goetmann. *Modern Portfolio Theory and Investment Analysis*. John Wiley & Sons, 2009.

- Lintner, John. 'The Valuation of Risk Assets and the Selection of Risky Investments in Stock Portfolios and Capital Budgets'. *Review of Economics and Statistics* 47, no. 1 (1965): 13–39.

- Sharpe, William F. '[Capital asset prices: A theory of market equilibrium under conditions of risk](#)'. *Journal of Finance* 19, no. 3 (1964): 425–442.

- Snowden, James. 'Should we give to more than one charity?' In *Effective Altruism: Philosophical Issues*, edited by Theron Pummer and Hilary Greaves. Oxford: Oxford University Press, 2019.

- Tran, Brigitte Roth. '[Divest, Disregard, or Double Down?](#)' *American Economic Review*: Insights 1, no. 2 (2019): 241–256.

Existing informal discussion:

- Kyle Bogosian, [Selecting investments based on covariance with the value of charities](#), 2017.

- Paul Christiano, [Giving now vs. later](#), 2013.

- John Halstead and Hauke Hillebrandt, [Impact investing is only a good idea in specific circumstances](#), 2018.

- Holden Karnofsky, [Worldview Diversification](#), 2016.

- Ben Kuhn, [How many causes should you give to?](#), 2014.

- Carl Shulman, [Salary or startup? How do-gooders can gain more from risky careers](#), 2012.

- Carl Shulman, [It's harder to favor a specific cause in more efficient charitable markets](#), 2014.

- Brian Tomasik, [When Should Altruists Be Financially Risk-Averse?](#), 2013.

2.5. Distributions of cost-effectiveness

Estimates of the effects of different interventions in different settings indicate that cost effectiveness can vary significantly, sometimes by multiple orders of magnitude, even within a given cause area. If so, this is important, because it pushes towards optimising for effectiveness over increasing the amount of resources going toward a cause. However, there is currently little rigorous investigation of the properties of the relevant cost-effectiveness distributions.

Economics:

- Establish more rigorously and more generally what can be said about typical distributions of cost-effectiveness, both within and between causes (Ord 2013)

(INFORMAL: Kaufman 2013; Kaufman 2015). How much of the variation of estimated cost effectiveness within a cause area is driven by differences in empirical settings or implementation (Meager 2020; Vivalt 2015)? How does variation of cost-effectiveness within a cause compare to variation of cost-effectiveness between causes (Vivalt 2020)? What are the implications for optimal diversification of cause areas and interventions?

ECON - PROGRAMME EVALUATION, EXTERNAL VALIDITY, BAYESIAN DECISION THEORY

- How does the estimated distribution of cost effectiveness affect the trade-off between the informational value of evaluating slightly different interventions in different settings versus the value created by implementing effective interventions given the existing state of knowledge (INFORMAL: Askell 2017)?

ECON - VALUE OF INFORMATION, BAYESIAN UPDATING

- How does the inclusion of indirect effects affect the estimated variance in cost-effectiveness across interventions?

ECON - PROGRAMME EVALUATION, ECONOMETRIC THEORY, STRUCTURAL MODELLING

Existing academic literature:

- Meager, Rachael. 'Aggregating Distributional Treatment Effects: A Bayesian Hierarchical Analysis of the Microcredit Literature'. Working paper, 2020.

- Ord, Toby. 'The Moral Imperative Towards Cost-Effectiveness in Global Health'. *Center for Global Development*, 2013.

- Tengs, Tammy O., Miriam E. Adams, Joseph S. Pliskin, Dana Gelb Safran, Joanna E. Siegel, Milton C. Weinstein and John D. Graham. 'Five-Hundred Life-Saving Interventions and Their Cost-Effectiveness'. *Risk Analysis* 15, no. 3 (1994): 369–390.

- Vivalt, Eva. 'Heterogeneous Treatment Effects in Impact Evaluation'. *American Economic Review, Papers and Proceedings* 105, no. 5 (2015): 467–470.

- ———. 'How Much Can We Generalize from Impact Evaluations?' *Journal of the European Economic Association*, 2020.

Existing informal discussion:

- Amanda Askell, The Moral Value of Information, 2017.

- Benjamin Todd, Is it fair to say that most social programmes don't work?, 2017.

- Holden Karnofsky, A conflict of Bayesian priors?, 2009.

- Jeff Kaufman, The Unintuitive Power Laws of Giving, 2013.

- Jeff Kaufman, Effectiveness: Gaussian?, 2015.

- Toby Ord, Taking charity seriously, 2013.

- Brian Tomasik, Why Charities Usually Don't Differ Astronomically in Expected Cost-Effectiveness, 2013.

2.6. Modelling altruism

Economic theory typically proceeds either (a) making minimally substantive assumptions about individuals' preferences (assuming only structural conditions, e.g., that preferences are complete and transitive), or (b) assuming that preferences are in some sense 'self-interested' (e.g., that an individual's utility depends only on his own consumption and leisure). Existing research shows that interesting new results can be established when we expand the domain of preferences to include the utility of others. However, this literature considers a relatively narrow domain of problems, and there is scope to further explore the implications of modelling agents as at least partially altruistic.

Economics:

- Are there settings in which agents have other-regarding preferences, and are either short-lived or have a non-zero discount rate, and are therefore unable to achieve a socially optimal outcome, for example, because they are unable to commit to 'punish' defectors to sustain an equilibrium (Rabin 1993; Povey 2014; Povey 2015)? What are the characteristics of these settings (Bolton and Ockenfels 2000)? Can we design mechanisms to overcome these challenges? Do these results have practical implications for decision-makers?

ECON - GAME THEORY, MECHANISM DESIGN

- How should we adapt key economic models to account for altruistic individuals with other-regarding preferences (Bergstrom 2002; Sobel 2005; Dufwenberg et al. 2011; Meager 2020)? Under what assumptions do key results, such as the Fundamental Theorems of Welfare Economics, still hold (Schall 1972; Pollack 1976; Rotemberg 2003)? In cases where they do not, can analogous results be derived?

ECON - MICROECONOMIC THEORY

- Is there a theoretical 'optimal' level of altruism in relevant settings (Povey 2015)? Do these results provide practical insights or implications for agents attempting to do good?

ECON - MICROECONOMIC THEORY, GAME THEORY

- Improve our understanding of the various motivations for apparently altruistic acts, for example, 'pure' altruism or 'warm glow' altruism (Andreoni 1990; Ashraf and Bandiera 2017). Which characteristics of individuals or the choices that they face are associated with different types of apparently altruistic acts?

ECON - BEHAVIOURAL ECONOMICS

Existing academic literature:

- Andreoni, James. 'Impure Altruism and Donations to Public Goods: A Theory of Warm-Glow Giving'. *The Economic Journal* 100, no. 401 (1990): 464–477.

- Ashraf, Nava and Oriana Bandiera. 'Altruistic Capital'. *American Economic Review* 107, no. 5 (2017): 70–75.

- Bergstrom, Theodore. 'Systems of Benevolent Utility Functions'. *Journal of Public Economic Theory* 1, no. 2 (1999): 71–100.

- Bolton, Gary E. and Axel Ockenfels. 'ERC: A Theory of Equity, Reciprocity, and Competition'. *American Economic Review* 90, no. 1 (2000): 166–193.

- Dufwenberg, Martin, Paul Heidhues, Georg Kirchsteiger, Frank Riedel and Joel Sobel. 'Other-Regarding Preferences in General Equilibrium'. *The Review of Economic Studies 78*, no. 2 (2011): 613–639.

- Fehr, Ernst and Klaus M. Schmidt. 'A Theory of Fairness, Competition, and Cooperation'. *Quarterly Journal of Economics* 114, no. 3 (1999): 817–868.

- Meager, Rachael. 'Aggregating Distributional Treatment Effects: A Bayesian Hierarchical Analysis of the Microcredit Literature'. Working paper, 2020.

- Pollack, Robert. 'Interdependent Preferences'. *American Economic Review* 66, no. 3 (1976): 309–320.

- Povey, Richard. 'The Limits to Altruism - A Survey'. Working paper, 2014.

- ———. 'The Socially Optimal Level of Altruism'. Working paper, 2015.

- Rabin, Matthew. 'Incorporating Fairness into Game Theory'. *American Economic Review* 83, no. 5 (1993): 1,281–1,302.

- Rotemberg, Julio J. 'The Benevolence of the Baker: Fair Pricing under Threat of Customer Anger'. Working paper, 2003.

- Schall, Lawrence D. 'Interdependent Utilities and Pareto Optimality'. *Quarterly Journal of Economics* 86, no. 1 (1972): 19–24.

- Sobel, Joel. 'Interdependent Preferences and Reciprocity'. *Journal of Economic Literature* 43, no. 2 (2005): 392–436.

2.7. Altruistic coordination

How should altruistically-motivated actors interact, particularly in cases where they have different values or beliefs regarding different philanthropic opportunities? For example, if two donors agree on the first-best use of money but disagree on the second-best, they each prefer that the other fully funds the first-best use. Similar coordination problems may arise when there is private information or comparative advantage of different actors contributing to different projects.

Economics:

- What are the implications of comparative advantage for a community of altruists, who may be heterogeneous in terms of resources, skills, information and values (INFORMAL: Todd 2018a)?

- How should game theoretic models be applied to analyse decisions faced by a community of altruists? For example, altruists with similar moral and empirical beliefs may face coordination problems similar to the 'stag hunt' game, whereby they can achieve a larger 'prize' if they coordinate, relative to working individually. (INFORMAL: Karnofsky 2014; Karnofsky 2015; Kuhn 2015; Ali and Karnofsky 2016; Cotton-Barratt and Leather 2016; Todd 2018b)

- How can results from the mechanism design literature help altruistic individuals and organisations to coordinate in a more effective manner (Andreoni 1998; Bracha et al. 2011; Conitzer and Sandholm 2011)? For example, how should agents with similar altruistic goals provide public goods (Goeree et al. 2005; Buterin et al. 2019)? In cases where individuals have (heterogeneous) private beliefs and/or information, which mechanisms can induce individuals to report these truthfully while respecting other desirable axioms (Peters 2019; Brandl et al. 2019)?

ECON - MECHANISM DESIGN, SOCIAL CHOICE THEORY

- Are there institutions or mechanisms that we can design to help improve the allocative efficiency of resources among altruists?

ECON - MECHANISM DESIGN

Interdisciplinary:

- How can a community of altruists with different moral and empirical views gain from trade? Do traditional challenges in trade extend to the case of moral trade (Ord 2015), for example, the Myerson-Satterthwaite theorem, according to which efficient trade cannot take place if two parties have private, stochastic valuations over the traded good? What are the challenges for moral trade that go beyond the challenges for ordinary trade? And can they be overcome (INFORMAL: Tomasik 2013; Oesterheld 2018)?

PHIL - MORAL UNCERTAINTY   ECON - GAME THEORY, MECHANISM DESIGN

Existing academic literature:

- Andreoni, James. 'Toward a Theory of Charitable Fund-Raising'. *Journal of Political Economy* 106, no. 6 (1998): 1,186–1,213.

- Bracha, Anat, Michael Menietti and Lise Vesterlund. 'Seeds to Succeed?' *Journal of Public Economics* 95, no. 5–6 (2011): 416–427.

- Brandl, Florian, Felix Brandt, Dominik Peters, Christian Stricker and Warut Suksompong. 'Donor coordination: Collective distribution of individual contributions'. Working paper, 2019.

- Buterin, Vitalik, Zoë Hitzig and E. Glen Weyl. 'A flexible design for funding public goods'. *Management Science* 65, no. 11 (2019): 5,171–5,187.

- Carpenter, Jeffrey, Jessica Holmes and Peter Hans Matthews. '"Bucket auctions" for charity'. *Games and Economic Behavior* 88 (2014): 260-276.

- Conitzer, Vincent and Tuomas Sandholm. 'Expressive Markets for Donating to Charities'. *Artificial Intelligence* 175, no. 7–8 (2011): 1,251–1,271.

- Goeree, Jacob K., Emiel Maasland, Sander Onderstal and John L. Turner. 'How (not) to raise money'. *Journal of Political Economy* 113, no. 4 (2005): 897–918.

- Myerson, Roger B. and Mark A. Satterthwaite. 'Efficient Mechanisms for Bilateral Trading'. *Journal of Economic Theory* 29, no. 2 (1983): 265–281.

- Ord, Toby. 'Moral Trade'. *Ethics* 126, no. 1 (2015): 118–138.

- Peters, Dominik. 'Economic Design for Effective Altruism'. In *The Future of Economic Design* by Jean-François Laslier, Hervé Moulin, M. Remzi Sanver, and William Zwicker. Cham, Switzerland : Springer, 2019.

Existing informal discussion:

- S. Nageeb Ali and Holden Karnofsky, [A conversation with Professor S. Nageeb Ali](#), 2016.

- Paul Christiano, [Certificates of Impact](#), 2014.

- Paul Christiano, [Repledge++](#), 2016.

- Owen Cotton-Barratt and Zachary Leather, [Donor coordination under simplifying assumptions](#), 2016.

- Max Dalton, [Mechanism design for altruistic cooperation](#), 2018.

- Holden Karnofsky, [Donor coordination and the "giver's dilemma"](#), 2014.

- Holden Karnofsky, [Good Ventures and giving now vs. later](#) (Section [Coordination issues](#)), 2015.

- Ben Kuhn, [Solving donation coordination problems](#), 2015.

- Rossa O'Keeffe-O'Donovan, [Economics of career choice: replaceability and comparative advantage](#), 2017.

- Caspar Oesterheld, [Multiverse-wide Cooperation via Correlated Decision Making](#), 2018.

- Michael Page, [Certificates of Impact](#), 2017.

- Carl Shulman, [Donor lotteries: demonstration and FAQ](#), 2016.

- Brian Tomasik, [Gains from Trade through Compromise](#), 2013.

- Benjamin Todd, [Should you play to your comparative advantage when choosing your career?](#), 2018a.

- Benjamin Todd, [Doing good together – how to coordinate effectively, and avoid single-player thinking](#), 2018b.

- Jess Whittlestone, [Building an Effective Altruism Community](#), 2017.

2.8. Institutions

Political and economic institutions have a tremendous influence on the lives of current and future generations (North 1990; Acemoglu et al. 2005). National governments and legislative bodies have access to vast resources and can shape formal institutions, such as constitutions, markets, laws and regulations, as well as informal constraints on human behaviour, such as social norms and culture (Alesina and Giuliano 2015). These institutions can also be influenced by other powerful groups, including corporations, lobbyists and non-governmental organisations. International institutions, including international organisations, treaties and agreements, can affect the interactions between nation states and other powerful entities. The effects of institutions can be large and persistent, so

determining which feasible reforms are likely to best promote social welfare may be highly valuable.

Philosophy:

- To what extent ought a government to take actions that are better for the world even if they conflict with the preferences of, and/or are worse for, their own citizens (Goodin 1995)? What about the relationship between corporate philanthropy and shareholder preference/interest?

PHIL - POLITICAL PHILOSOPHY, DUTIES OF BENEFICENCE, BUSINESS ETHICS

- What can considerations of international distributive justice (Blake and Smith 2020) tell us about a fair allocation of benefits and burdens related to institutional mechanisms that facilitate international cooperation in the provision of global public goods, such as the mitigation of global catastrophic risks from emerging technologies (Dafoe 2018)?

PHIL - POLITICAL PHILOSOPHY

Economics:

- What is the optimal design of international institutions that are formed to increase global public goods or decrease global public bads (Buchholz and Sandler forthcoming)? Can institutions be designed to overcome the participation and incentive compatibility constraints of potentially self-interested nation states, while achieving globally socially optimal investments? Can such mechanisms be enforced?

ECON - GAME THEORY, MECHANISM DESIGN, POLITICAL ECONOMY

- Which types of political systems and institutions place a higher weight on long-run outcomes? For example, are democratic systems fundamentally more 'short-termist' than non-democratic systems, because of the political pressure of elections (Harstad 2020)? If yes, to what extent can such short-termist pressures be mitigated?

ECON - POLITICAL ECONOMY

- Some political decisions have persistent and widespread long-run effects, for example, drafting or amending constitutions, extending suffrage or political rights, or enshrining human rights. At certain points in time, making such changes may be particularly tractable. How do political institutions evolve over time (Acemoglu et al. 2015)? To what extent can the tractability and persistence of policies be predicted and leveraged (Coate and Morris 1999)? How tractable are changes to institutions that may increase long-run growth? How tractable are changes to institutions that may have other long-run benefits, for example, strengthening their resilience to catastrophic risks?

ECON - POLITICAL ECONOMY, FORECASTING, ECONOMIC GROWTH

Interdisciplinary:

- Should institutional actors with access to a large amount of resources assess expected value, risk and/or diversification in a different way to individuals when evaluating opportunities to do good (Kagan 2011; McMahan 2016; Kissel 2017; Collins 2019) (INFORMAL: Karnofsky 2013; Reich 2015; Greaves 2017)?

PHIL - DECISION THEORY   ECON - DECISION THEORY, MODERN PORTFOLIO THEORY

- Most of the individuals who are impacted by government decisions are people in the future or non-human animals. They do not get a vote, nor do they participate in markets. To what extent does this provide an argument against statist political philosophies, perhaps analogous to the ways in which market failures justify deviations from a free market? Can we design policies, mechanisms or institutions that overcome these failures (Rawls 1971; Barry 1997; Donaldson and Kymlicka 2013)?

PHIL - POLITICAL PHILOSOPHY   ECON - POLITICAL ECONOMY, SOCIAL CHOICE THEORY

- What is the optimal level of diversity along various dimensions (and, relatedly, the optimal level of freedom for individual institutions to go their own ways) that appropriately balances upsides of diversity, like information value and resiliency to certain kinds of risk, against downsides of diversity, like prima facie inefficiencies from forgone moral trade and the "unilateralist's curse" (Bostrom et al. 2016)?

PHIL - POLITICAL PHILOSOPHY   ECON - POLITICAL ECONOMY, VALUE OF INFORMATION

- Some researchers distinguish between 'culture' and institutions, defining culture as "beliefs and values that ethnic, religious, and social groups transmit fairly unchanged from generation to generation" (Guiso et al. 2006). What is the two-way causal relationship between culture and institutions (Alesina and Giuliano 2015)? Are changes to cultural traits that may promote social welfare, such as "generalised morality" (Platteau 2000), persistent and tractable to promote? Should we expect evolutionary pressures to favour certain values or beliefs over time (Buchanan and Powell 2018; Buchanan 2020)?

PHIL - POLITICAL PHILOSOPHY   ECON - ECONOMICS OF CULTURE

Existing academic literature:

- Acemoglu, Daron, Georgy Egorov and Konstantin Sonin. 'Political Economy in a Changing World'. *Journal of Political Economy* 123, no. 5 (2015): 1,038–1,086.

- Alesina, Alberto and Paola Giuliano. 'Culture and Institutions'. *Journal of Economic Literature* 53, no. 4: 898–944, 2015.

- Barry, Brian. 'Sustainability and Intergenerational Justice'. *Theoria* 44, no. 89 (1997): 43–64.

- Blake, Michael and Patrick TaylorSmith. 'International Distributive Justice'. In *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta, 2020.

- Bostrom, Nick, Thomas Douglas and Anders Sandberg. 'The Unilateralist's Curse and the Case for a Principle of Conformity'. *Social Epistemology* 30, no. 4 (2016): 350–371.

- Buchanan, Allen. *Our Moral Fate-Evolution and the Escape from Tribalism*. Cambridge, MA: MIT Press, 2020.

- Buchanan, Allen and Russell Powell. *The evolution of moral progress: a biocultural theory*. Oxford: Oxford University Press, 2018.

- Buchholz, Wolfgang and Todd Sandler. 'Global Public Goods: A Survey'. *Journal of Economic Literature*, forthcoming.

- Coate, Stephen, and Stephen Morris. 'Policy persistence'. *American Economic Review* 89, no. 5 (1999): 1,327–1,336.

- Collins, Stephanie. 'Beyond Individualism'. In *Effective Altruism: Philosophical Issues*, edited by Hilary Greaves and Theron Pummer. Oxford: Oxford University Press, 2019.

- Dafoe, Allan. 'AI Governance: A Research Agenda'. *Center for the Governance of AI, Future of Humanity Institute*, 2018.

- Donaldson, Sue and Will Kymlicka. *Zoopolis: A Political Theory of Animal Rights*. Oxford; New York: Oxford University Press, 2013.

- Goodin, Robert E. *Utilitarianism as a Public Philosophy*. Cambridge, UK: Cambridge University Press, 1995.

- Guiso, Luigi, Paola Sapienza and Luigi Zingales. 'Does Culture Affect Economic Outcomes?' *Journal of Economic Perspectives* 20, no. 2 (2006): 23–48.

- Hanson, Robin. 'Shall We Vote on Values, But Bet on Beliefs?' *Journal of Political Philosophy* 21, no. 2 (2013): 151–178.

- Harstad, Bård. 'Technology and time inconsistency'. *Journal of Political Economy* 128, no. 7 (2020): 2,653–2,689.

- Kagan, Shelly. 'Do I Make a Difference?'. *Philosophy & Public Affairs* 39, no. 2 (2011): 105–141.

- Kissel, Joshua. 'Effective Altruism and Anti-Capitalism: An Attempt at Reconciliation'. *Essays in Philosophy* 18, no. 1 (2017): 1–23.

- Lalley, Steven P. and E. Glen Weyl. 'Quadratic Voting: How Mechanism Design Can Radicalize Democracy'. *AEA Papers and Proceedings* 108 (2018): 33–37.

- Maskin, Eric S. 'Mechanism Design: How to Implement Social Goals'. *American Economic Review* 98, no. 3 (2008): 567–576.

- McMahan, Jeff. 'Philosophical Critiques of Effective Altruism'. *The Philosophers' Magazine*, no. 73 (2016): 92–99.

- North, Douglass. *Institutions, institutional change and economic performance*. New York: Cambridge University Press, 1990.

- Platteau, Jean-Philippe. *Institutions, social norms, and economic development*. Vol. 1. London: Psychology Press, 2000.

- Rawls, John. *A Theory of Justice*. Cambridge, MA: Belknap Press, 1971.

Existing informal discussion:

- Scott Alexander, Beware Systemic Change, 2015.

- Hilary Greaves, The collectivist critique of the effective altruist movement,2017.

- The Open Philanthropy Project, U.S. Policy, 2018.

- Holden Karnofsky, The Role of Philanthropic Funding in Politics, 2013.

- Rob Reich, The Logic of Effective Altruism, 2015.

Bibliography

Acemoglu, Daron, Georgy Egorov and Konstantin Sonin. 'Political Economy in a Changing World'. *Journal of Political Economy* 123, no. 5 (2015): 1,038–1,086.

Acemoglu, Daron, Simon Johnson and James A. Robinson. 'Institutions as a Fundamental Cause of Long-Run Growth'. *Handbook of Economic Growth*, 1A (2005): 385–472.

Aghion, Philippe, Benjamin F. Jones and Charles I. Jones. 'Artificial intelligence and Economic Growth'. No. w23928. National Bureau of Economic Research, 2017.

Alesina, Alberto and Giuliano, Paola (2015). 'Culture and Institutions'. *Journal of Economic Literature* 53, no. 4: 898–944.

Andreoni, James. 'Impure Altruism and Donations to Public Goods: A Theory of Warm-Glow Giving'. *The Economic Journal* 100, no. 401 (1990): 464–477.

———. 'Toward a Theory of Charitable Fund-Raising'. *Journal of Political Economy* 106, no. 6 (1998): 1,186–1,213.

Arntzenius, Frank. 'Utilitarianism, Decision Theory, and Eternity'. *Philosophical Perspectives* 28, no. 1 (2014): 31–58.

Arrhenius, Gustaf. *Population Ethics: the Challenge of Future Generations*. Unpublished manuscript.

Aschenbrenner, Leopold. 'Existential Risk and Growth'. Working paper, 2020.

Asheim, Geir B. 'Intergenerational Equity'. *Annual Review of Economics* 2, no. 1 (2010): 197–222.

Ashraf, Nava, and Oriana Bandiera. 'Altruistic Capital'. *American Economic Review* 107, no. 5 (2017): 70–75.

Askell, Amanda. 'Evidence Neutrality and the Moral Value of Information'. In *Effective Altruism: Philosophical Issues*, edited by Hilary Greaves and Theron Pummer. Oxford: Oxford University Press, 2019.

Athey, Susan, Raj Chetty, Guido Imbens, Hyunseung Kang. 'The Surrogate Index: Combining Short-term Proxies to Estimate Long-term Treatment Effects More Rapidly and Precisely'. *Randomized Social Experiments Ejournal*, 2003.

Baranzini, Andrea, and François Bourguignon. 'Is Sustainable Growth Optimal?' *International Tax and Public Finance* 2.2 (1995): 341–356.

Barrage, Lint. 'Be Careful What You Calibrate for: Social Discounting in General Equilibrium'. *Journal of Public Economics* 160 (2018): 33–49.

Barrett, Jacob. 'Social Reform in a Complex World'. *Journal of Ethics and Social Philosophy* 17, no. 2 (2020): 103–132.

Barry, Brian. 'Sustainability and Intergenerational Justice'. *Theoria* 44, no. 89 (1997): 43–64.

Basu, Kaushik, and Tapan Mitra. 'Aggregating Infinite Utility Streams with Intergenerational Equity: The Impossibility of Being Paretian'. *Econometrica* 71, no. 5 (2003): 1,557–1,563.

Baum, Seth D., Stuart Armstrong, Timoteus Ekenstedt, Olle Häggström, Robin Hanson, Karin Kuhlemann, Matthijs M. Maas et al. 'Long-Term Trajectories of Human Civilization'. *Foresight*, 2019.

Beckman, Ludvig. *The Frontiers of Democracy: The Right to Vote and Its Limits*. Basingstoke: Palgrave MacMillan, 2009.

Beckstead, Nicholas. 'On the Overwhelming Importance of Shaping the Far Future'. PhD dissertation. New Brunswick: Rutgers University, 2013.

———. 'A Brief Argument for the Overwhelming Importance of Shaping the Far Future'. In *Effective Altruism: Philosophical Issues*, edited by Hilary Greaves and Theron Pummer. Oxford: Oxford University Press, 2019.

Bergstrom, Theodore. 'Systems of Benevolent Utility Functions'. *Journal of Public Economic Theory* 1, no. 2 (1999): 71–100.

Bishop, Richard C. 'Option Value: An Exposition and Extension'. *Land Economics* 58, no. 1 (1982): 1–15.

Blackorby, Charles, Walter Bossert, and David Donaldson. 'Intertemporal population ethics: critical-level utilitarian principles'. *Econometrica: Journal of the Econometric Society* (1995): 1,303–1,320.

Blake, Michael and Patrick Taylor Smith. 'International Distributive Justice'. In *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta, 2020.

Bolton, Gary E and Axel Ockenfels. 'ERC: A Theory of Equity, Reciprocity, and Competition'. *American Economic Review* 90, no. 1 (2000): 166–193.

Bommier, Antoine, Bruno Lanz and Stéphane Zuber. 'Models-as-usual for unusual risks? On the value of catastrophic climate change.' *Journal of Environmental Economics and Management* 74 (2015): 1–22.

Bostrom, Nick. 'Existential Risks: Analyzing Human Extinction Scenarios and Related Hazards'. *Journal of Evolution and Technology* 9 (2002).

———. 'Astronomical Waste: The Opportunity Cost of Delayed Technological Development'. *Utilitas* 15, no. 3 (2003): 308–314.

———. 'What is a Singleton?' *Linguistic and Philosophical Investigations* 5, no. 2 (2006): 48–54.

———. 'Technological Revolutions: Ethics and Policy in the Dark'. In *Nanoscale*, edited by Nigel M. de S. Cameron and M. Ellen Mitchell. Hoboken, NJ: John Wiley & Sons, 2007: 129–152.

———. 'Pascal's Mugging'. *Analysis* 69, no. 3 (2009): 443–445.

———. 'Infinite Ethics'. *Analysis and Metaphysics*, no. 10 (2011): 9–59.

———. 'Existential Risk Prevention as Global Priority'. *Global Policy* 4, no. 1 (2013): 15–31.

———. *Superintelligence: Paths, Dangers, Strategies*. First edition. Oxford: Oxford University Press, 2014.

Bostrom, Nick, Thomas Douglas and Anders Sandberg. 'The Unilateralist's Curse and the Case for a Principle of Conformity'. *Social Epistemology* 30, no. 4 (2016): 350–371.

Bouguen, Adrien, Yue Huang, Michael Kremer and Edward Miguel. 'Using RCTs to Estimate Long-Run Impacts in Development Economics'. No. w25356. National Bureau of Economic Research, 2018.

Bracha, Anat, Michael Menietti and Lise Vesterlund. 'Seeds to Succeed?' *Journal of Public Economics* 95, no. 5–6 (2011): 416–427.

Brandl, Florian, Felix Brandt, Dominik Peters, Christian Stricker and Warut Suksompong. 'Donor coordination: Collective distribution of individual contributions'. Working paper, 2019.

Brock, Dan W. 'Separate Spheres and Indirect Benefits'. *Cost Effectiveness and Resource Allocation* 1, no. 4 (2003).

Broome, John. *Weighing Lives*. Oxford; New York: Oxford University Press, 2004.

Buchak, Lara. *Risk and Rationality*. Oxford: Oxford University Press, 2013.

Buchanan, Allen. *Our Moral Fate-Evolution and the Escape from Tribalism*. Cambridge, MA: MIT Press, 2020.

Buchanan, Allen and Russell Powell. *The evolution of moral progress: a biocultural theory*. Oxford: Oxford University Press, 2018.

Buchholz, Wolfgang and Todd Sandler. 'Global Public Goods: A Survey'. *Journal of Economic Literature*, forthcoming.

Buterin, Vitalik, Zoë Hitzig and E. Glen Weyl. 'A flexible design for funding public goods'. *Management Science* 65, no. 11 (2019): 5,171–5,187.

Camerer, Colin F., Anna Dreber, Eskil Forsell, Teck-Hua Ho, Jürgen Huber, Magnus Johannesson, Michael Kirchler, Johan Almenberg, Adam Altmejd, Taizan Chan, Emma Heikensten, Felix Holzmeister, Taisuke Imai, Siri Isaksson, Gideon Nave, Thomas Pfeiffer, Michael Razen and Hang Wu 'Evaluating replicability of laboratory experiments in economics'. *Science* 351.6280 (2016): 1,433–1,436.

Carpenter, Jeffrey, Jessica Holmes, and Peter Hans Matthews. '"Bucket auctions" for charity'. *Games and Economic Behavior* 88 (2014): 260–276.

Christensen, David. 'Epistemology of Disagreement: The Good News'. *Philosophical Review* 116, no. 2 (2007): 187–217.

———. 'Disagreement as Evidence: The Epistemology of Controversy'. *Philosophy Compass* 4, no. 5 (2009): 756–767.

Christensen, David and Jennifer Lackey, eds. *The Epistemology of Disagreement: New Essays*. Oxford: Oxford University Press, 2013.

Christensen, Peter, Kenneth Gillingham and William Nordhaus. 'Uncertainty in forecasts of long-run economic growth.' *Proceedings of the National Academy of Sciences* 115.21 (2018): 5,409–5,414.

Coate, Stephen and Stephen Morris. 'Policy persistence'. *American Economic Review* 89, no. 5 (1999): 1,327–1,336.

Cockburn, John, Jean-Yves Duclos and Agnès Zabsonré. 'Is global social welfare increasing? A critical-level enquiry'. *Journal of Public Economics* 118 (2014): 151–162.

Collins, Stephanie. 'Beyond Individualism'. In *Effective Altruism: Philosophical Issues*, edited by Hilary Greaves and Theron Pummer. Oxford: Oxford University Press, 2019.

Conitzer, Vincent and Tuomas Sandholm. 'Expressive Markets for Donating to Charities'. *Artificial Intelligence* 175, no. 7–8 (2011): 1,251–1,271.

Cotton-Barratt, Owen. 'Allocating Risk Mitigation across Time'. *Future of Humanity Institute*, Technical Report #2015-2.

Cotton-Barratt, Owen and Toby Ord. 'Existential Risk and Existential Hope: Definitions'. *Future of Humanity Institute*, Technical Report #2015-1.

Cowen, Tyler. 'Caring About the Distant Future: Why It Matters and What It Means'. *The University Of Chicago Law Review* 74, no. 5 (2007): 5–40.

———. *Stubborn Attachments*. San Francisco: Stripe Press, 2018.

Cowie, Christopher. 'Does the repugnant conclusion have any probative force?'. *Philosophical Studies* 174.12 (2017): 3,021–3,039.

Dafoe, Allan. 'AI Governance: A Research Agenda'. *Center for the Governance of AI, Future of Humanity Institute*, 2018.

Dasgupta, Partha. 'Discounting climate change.' *Journal of Risk and Uncertainty* 37.2-3 (2008): 141–169.

DellaVigna, Stefano and Devin Pope. 'Predicting experimental results: who knows what?' *Journal of Political Economy* 126.6 (2018): 2,410–2,456.

Deutsch, David. *The beginning of infinity: Explanations that transform the world*. Penguin UK, 2011.

Dixit, Avinash and Robert S. Pindyck. *Investment Under Uncertainty*. Princeton: Princeton University Press, 1994.

Donaldson, Sue and Will Kymlicka. *Zoopolis: A Political Theory of Animal Rights*. Oxford; New York: Oxford University Press, 2013.

Dreber, Anna, Thomas Pfeiffer, Johan Almenberg, Siri Isaksson, Brad Wilson, Yiling Chen, Brian A. Nosek, and Magnus Johannesson. 'Using prediction markets to estimate the reproducibility of scientific research'. *Proceedings of the National Academy of Sciences* 112.50 (2015): 15,343–15,347.

Dufwenberg, Martin, Paul Heidhues, Georg Kirchsteiger, Frank Riedel, and Joel Sobel. 'Other-Regarding Preferences in General Equilibrium'. *The Review of Economic Studies* 78, no. 2 (2011): 613–639.

Easterlin, Richard A. 'Paradox lost?.' USC-INET Research Paper 16-02 (2016).

Elga, Adam. 'Reflection and Disagreement'. *Noûs* 41, no. 3 (2007): 478–502.

Elton, Edwin, Martin Gruber, Stephen Brown and William Goetmann. *Modern Portfolio Theory and Investment Analysis*. John Wiley & Sons, 2009.

Emerson, Jed, Jay Wachowicz and Suzi Chun. 'Social Return on Investment: Exploring Aspects of Value Creation in the Nonprofit Sector'. *Social Purpose Enterprises and Venture Philanthropy in the New Millennium* 2 (2000): 132–173.

Erev, Ido, Eyal Ert, Alvin E. Roth, Ernan Haruvy, Stefan M. Herzog, Robin Hau, Ralph Hertwig, Terrence Stewart, Robert West and Christian Lebiere. 'A choice prediction competition: Choices from experience and from description'. *Journal of Behavioral Decision Making* 23.1 (2010): 15–47.

Estlund, David. *Utopophobia: On the Limits (If Any) of Political Philosophy*. Princeton, NJ: Princeton University Press, 2019.

Fehr, Ernst and Klaus M. Schmidt. 'A Theory of Fairness, Competition, and Cooperation'. *Quarterly Journal of Economics* 114, no. 3 (1999): 817–868.

Feldman, Richard and Ted A. Warfield, eds. *Disagreement*. Oxford: Oxford University Press, 2010.

Fleurbaey, Marc and Stéphane Zuber. 'Discounting, Risk and Inequality: A General Approach'. *Journal of Public Economics* 128 (2015): 34–49.

Fleurbaey, Marc, Maddalena Ferranna, Mark Budolfson, Francis Dennig, Kian Mintz-Woo, Robert Socolow, Dean Spears and Stéphane Zuber. 'The Social Cost of Carbon: Valuing Inequality, Risk, and Population for Climate Policy'. *The Monist* 102.1 (2019): 84–109.

Frank, Joshua. 'Is there an "animal welfare Kuznets curve"?.' *Ecological Economics* 66.2-3 (2008): 478–491.

Friedman, Benjamin M. *The Moral Consequences of Economic Growth*. New York: Vintage Books, 2006.

Frumkin, Peter. *Strategic Giving: The Art and Science of Philanthropy*. Chicago: University of Chicago Press, 2016.

Gaus, Gerald. *The Tyranny of the Ideal: Justice in a Diverse Society*. Princeton, NJ: Princeton University Press, 2016.

Gibbard, Allan. 'Manipulation of Voting Schemes: A General Result'. *Econometrica* 41, no. 4 (1973): 587–601.

Goeree, Jacob K., Emiel Maasland, Sander Onderstal, and John L. Turner. 'How (not) to raise money'. *Journal of Political Economy* 113, no. 4 (2005): 897–918.

Gollier, Christian. 'Maximizing the Expected Net Future Value as an Alternative Strategy to Gamma Discounting'. *Finance Research Letters* 1, no. 2 (2004): 85–89.

Gonzalez-Ricoy, Inigo and Axel Gosseries, eds. *Institutions for Future Generations*. Oxford: Oxford University Press, 2017.

Goodin, Robert E. *Utilitarianism as a Public Philosophy*. Cambridge, UK: Cambridge University Press, 1995.

———. 'Enfranchising All Affected Interests, And Its Alternatives'. *Philosophy and Public Affairs* 35, no. 1 (2007): 40–68.

Gordon, Robert J. 'The Demise of US Economic Growth: Restatement, Rebuttal, and Reflections'. No. w19895. National Bureau of Economic Research, 2014.

Granger, Clive WJ and Yongil Jeon. 'Long-term forecasting and evaluation.' *International Journal of Forecasting* 23.4 (2007): 539–551.

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

———. 'Discounting for Public Policy: A Survey'. *Economics & Philosophy* 33, no. 3 (2017): 391–439.

———. 'Climate Change and Optimum Population'. *The Monist* 102, no. 1 (2019): 42–65.

———. 'Discounting Future Health'. In *Global Health Priority-Setting: Cost-Effectiveness and Beyond*, edited by Ruger and Verguet Otterson Millum Johansson Jamison Emanuel Norheim. Oxford: Oxford University Press, forthcoming.

———. 'Optimum Population Size'. In *Oxford Handbook of Population Ethics*, edited by Gustaf Arrhenius, Krister Bykvist and Tim Campbell. Oxford: Oxford University Press, forthcoming.

Greaves, Hilary and Owen Cotton-Barratt. 'A Bargaining-theoretic approach to moral uncertainty'. Working paper, 2019.

Greaves, Hilary and Toby Ord. 'Moral Uncertainty About Population Axiology'. *Journal of Ethics & Social Philosophy* 12, no. 2 (2017): 135–167.

Greaves, Hilary, Andreas L. Mogensen and William MacAskill. 'Longtermism for Risk Averse Altruists'. Manuscript in preparation.

Guiso, Luigi, Paola Sapienza and Luigi Zingales. 'Does Culture Affect Economic Outcomes?' *Journal of Economic Perspectives* 20, no. 2 (2006): 23–48.

Gustafsson, Johan E., and Martin Peterson. 'A computer simulation of the argument from disagreement.' *Synthese* 184.3 (2012): 387–405.

Hanson, Robin. 'Long-term growth as a sequence of exponential modes'. Working manuscript, 2000.

———. 'Economic growth given machine intelligence'. Technical Report, University of California, Berkeley, 2001.

———. 'Shall We Vote on Values, But Bet on Beliefs?' *Journal of Political Philosophy* 21, no. 2 (2013): 151–178.

Harstad, Bård. 'Technology and time inconsistency'. *Journal of Political Economy* 128, no. 7 (2020): 2,653–2,689.

Hurka, Thomas. 'Asymmetries In Value'. *Noûs* 44, no. 2 (2010): 199–223.

Huth, Paul K. and Todd L. Allee. *The Democratic Peace and Territorial Conflict in the Twentieth Century*. Vol. 82. Cambridge University Press, 2002.

IPCC. *Climate change 2014: Impacts, adaptation, and vulnerability. Part A: Global and sectoral aspects. Contribution of Working Group II to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change.* Cambridge, UK and New York, NY, USA: Cambridge University Press, 2014.

Irvin, Renée A. 'Endowments: Stable Largesse or Distortion of the Polity?' *Public Administration Review* 67, no. 3 (2007): 445–457.

Isaacs, Yoaav. 'Probabilities cannot be rationally neglected'. *Mind* 125, no. 499 (2016) 759–762.

Jaakkola, Niko and Antony Millner. 'Non-dogmatic climate policy'. Working paper, 2020.

Jansen, Christoph, Georg Schollmeyer and Thomas Augustin. 'Concepts for Decision Making under Severe Uncertainty with Partial Ordinal and Partial Cardinal Preferences'. *International Journal of Approximate Reasoning* 98 (2018): 112–131.

Jansen, Paul and David Katz. 'For Nonprofits, Time Is Money'. *The McKinsey Quarterly* 1 (2002): 124–133.

Jones, Charles I. 'Life and Growth'. *Journal of Political Economy* 124, no. 2 (2016): 539–578.

———. 'The Facts of Economic Growth'. *Handbook of Macroeconomics*, 2A (2016): 3–69.

———. 'The End of Economic Growth? Unintended Consequences of a Declining Population'. Working paper, 2020.

Kagan, Shelly. 'Do I Make a Difference?' *Philosophy & Public Affairs* 39, no. 2 (2011): 105–141.

Kamm, Frances M. *Morality, Mortality; Volume I: Death and Whom to Save From It*. Oxford: Oxford University Press, 1998.

Kavka, Gregory and Virginia Warren. 'Political Representation for Future Generations'. In *Environmental Philosophy: A Collection of Readings*, edited by Robert Elliott and Arran Gare. University Park, PA: the Pennsylvania State University Press, 1983.

Kimball, Miles S. 'Making Sense of Two-Sided Altruism'. *Journal of Monetary Economics* 20, no. 2 (1987): 301–326.

Kissel, Joshua. 'Effective Altruism and Anti-Capitalism: An Attempt at Reconciliation'. *Essays in Philosophy* 18, no. 1 (2017): 1–23.

Klausner, Michael D. 'When Time Isn't Money: Foundation Payouts and the Time Value of Money'. *SSRN Electronic Journal*, 2003.

Klenow, Peter J., and Chad I. Jones. 'Beyond GDP? Welfare across Countries and Time'. *American Economic Review* 106, no. 9 (2016): 2,426–2,457.

Kott, Alexander and Philip Perconti. 'Long-term forecasts of military technologies for a 20–30 year horizon: An empirical assessment of accuracy'. *Technological Forecasting and Social Change* 137 (2018): 272–279.

Lalley, Steven P. and E. Glen Weyl. 'Quadratic Voting: How Mechanism Design Can Radicalize Democracy'. *AEA Papers and Proceedings* 108 (2018): 33–37.

Landesman, Cliff. 'When to Terminate a Charitable Trust?' *Analysis* 55, no. 1 (1995): 12–13.

Lauwers, Luc and Peter Vallentyne. 'Infinite Utilitarianism: More Is Always Better'. *Economics & Philosophy* 20, no. 2 (2004): 307–330.

Lawford-Smith, Holly. 'Understanding Political Feasibility'. *Journal of Political Philosophy* 21, no. 3 (2013): 243–259.

Lempert, Robert J. Popper, Steven W. and Bankes, Steven C. 'Shaping the next one hundred years: new methods for quantitative, long-term policy analysis.' RAND Corporation, 2003.

Lenman, James. 'Consequentialism and Cluelessness'. *Philosophy and Public Affairs* 29, no. 4 (2000): 342–370.

Li, Shengwu. 'Obviously Strategy-Proof Mechanisms'. *American Economic Review* 107, no. 11 (2017): 3,257–3,287.

Lintner, John. 'The Valuation of Risk Assets and the Selection of Risky Investments in Stock Portfolios and Capital Budgets'. *Review of Economics and Statistics* 47, no. 1 (1965): 13–39.

Lippert-Rasmussen, Kasper and Sigurd Lauridsen. 'Justice and the Allocation of Healthcare Resources: Should Indirect, Non-Health Effects Count?' *Medicine, Health Care and Philosophy* 13, no. 3 (2010): 237–246.

MacAskill, William. 'Practical Ethics Given Moral Uncertainty'. *Utilitas* 31, no. 3 (2019): 231–245.

———. 'Human Extinction, Asymmetry, and Option Value'. Manuscript in preparation.

———. 'When Should an Effective Altruist Donate?' Working paper, 2019.

MacAskill, William, Krister Bykvist and Toby Ord. *Moral Uncertainty*. Oxford University Press, 2020.

MacAskill, William and Toby Ord. 'Why Maximize Expected Choice-Worthiness?' *Noûs*, 14 July 2018.

Martin, Ian W. R., and Robert S. Pindyck. 'Averting Catastrophes: The Strange Economics of Scylla and Charybdis'. *American Economic Review* 105, no. 10 (2015): 2,947–2,985.

———. 'Averting Catastrophes that Kill'. Working paper, 2017.

Maskin, Eric S. 'Mechanism Design: How to Implement Social Goals'. *American Economic Review* 98, no. 3 (2008): 567–76.

Matheny, Gaverick and Kai Chan. 'Human Diets and Animal Welfare: The Illogic of the Larder'. *Journal of Agricultural and Environmental Ethics* 18, no. 6 (2005): 57–94.

Matheny, Jason G. 'Reducing the Risk of Human Extinction'. *Risk Analysis* 27, no. 5 (2007): 1,335–1,344.

McMahan, Jeff. 'Philosophical Critiques of Effective Altruism'. *The Philosophers' Magazine*, no. 73 (2016): 92–99.

Meager, Rachael. 'Aggregating Distributional Treatment Effects: A Bayesian Hierarchical Analysis of the Microcredit Literature'. Working paper, 2020.

Méjean, Aurélie, Antonin Pottier, Stéphane Zuber and Marc Fleurbaey. 'Intergenerational Equity under Catastrophic Climate Change'. Working paper, 2017.

Millner, Antony. 'On Welfare Frameworks and Catastrophic Climate Risks'. *Journal of Environmental Economics and Management* 65, no. 2 (2013): 310–325.

———. 'Nondogmatic Social Discounting'. *American Economic Review* 110.3 (2020): 760–75.

Mogensen, Andreas L. '"The only ethical argument for positive δ"?' Working paper, 2019.

———. 'Meaning, Medicine, and Merit'. *Utilitas* 32, no.1 (2020), 90–107

———. 'Maximal cluelessness'. *The Philosophical Quarterly* (2020).

———. 'Moral Demands and the Far Future'. *Philosophy and Phenomenological Research*, 2020.

Moller, Dan. 'Should We Let People Starve - for Now?' *Analysis* 66, no. 3 (2006): 240–247.

Monton, Bradley. 'How to avoid maximizing expected utility'. *Philosophers' Imprint* 19 (2019): 1–25.

Mortensen, Dale. 'Job Search and Labor Market Analysis'. *Handbook of Labor Economics* 2 (1986): 849–919.

Müller, Ulrich K. and Mark W. Watson. 'Measuring uncertainty about long-run predictions.' *Review of Economic Studies* 83.4 (2016): 1,711–1,740.

Myerson, Roger B. 'Incentive Compatibility and the Bargaining Problem'. *Econometrica* 47, no. 1 (1979): 61–73.

Myerson, Roger B. and Mark A. Satterthwaite. 'Efficient Mechanisms for Bilateral Trading'. *Journal of Economic Theory* 29, no. 2 (1983): 265–281.

National Research Council. 'The Accuracy of Past Projections.' Ch. 2 in: *Beyond six billion: Forecasting the world's population*. National Academies Press, 2000.

Ng, Yew-Kwang. 'Social criteria for evaluating population change: an alternative to the Blackorby-Donaldson criterion'. *Journal of Public Economics* 29.3 (1986): 375–381.

———. 'What Should We Do About Future Generations? Impossibility of Parfit's Theory X'. *Economics and Philosophy* 5, no. 2 (1989): 235–253.

———. 'Should we be very cautious or extremely cautious on measures that may involve our destruction?' *Social Choice and Welfare* 8, no. 1 (1991): 79–88.

———. 'The Importance of Global Extinction in Climate Change Policy'. *Global Policy* 7, no. 3 (2016): 315–322.

Nordhaus, William D. 'A Review of the Stern Review on the Economics of Climate Change'. *Journal of Economic Literature* 45, no. 3 (2007): 686–702.

———. 'The economics of tail events with an application to climate change'. *Review of Environmental Economics and Policy* 5.2 (2011): 240–257.

———. 'Are We Approaching an Economic Singularity? Information Technology and the Future of Economic Growth'. Cowles Foundation Discussion Paper 2021, 2015.

North, Douglass. *Institutions, institutional change and economic performance*. New York: Cambridge University Press, 1990.

Olsson, Caroline and Gustav Alexandrie. 'A Beef with Growth? An Empirical Analysis of Income and Animal Farming'. Bachelor Thesis, Stockholm University 2019.

Ord, Toby. 'The Moral Imperative Towards Cost-Effectiveness in Global Health'. *Center for Global Development*, 2013.

———. 'Moral Trade'. *Ethics* 126, no. 1 (2015): 118–138.

———. *The Precipice: Existential Risk and the Future of Humanity*. London: Bloomsbury, 2020.

Parfit, Derek. *Reasons and Persons*. Oxford: Oxford University Press, 1984.

———. *On What Matters*. Oxford University Press, 2011.

Peters, Dominik. 'Economic Design for Effective Altruism'. In *The Future of Economic Design* by Jean-François Laslier, Hervé Moulin, M. Remzi Sanver, and William Zwicker. Cham, Switzerland: Springer, 2019.

Peterson, Martin. 'The St. Petersburg Paradox'. In *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta, 2020.

Platteau, Jean-Philippe. *Institutions, social norms, and economic development*. Vol. 1. London: Psychology Press, 2000.

Prelec, Dražen. 'A Bayesian truth serum for subjective data'. *Science* 306, no. 5,695 (2004): 462–466.

Pollack, Robert. 'Interdependent Preferences'. *American Economic Review* 66, no. 3 (1976): 309–320.

Povey, Richard. 'The Limits to Altruism - A Survey'. Working paper, 2014.

———. 'The Socially Optimal Level of Altruism'. Working paper, 2015.

Quiggin, John. 'A Theory of Anticipated Utility'. *Journal of Economic Behavior & Organization* 3, no. 4 (1982): 323–343.

Rabin, Matthew. 'Incorporating Fairness into Game Theory'. *American Economic Review* 83, no. 5 (1993): 1,281–1,302.

Ramsey, Frank Plumpton. 'A Mathematical Theory of Saving'. *The Economic Journal* 38.152 (1928): 543–559

Rawls, John. *A Theory of Justice*. Cambridge, MA: Belknap Press, 1971.

Risi, Joseph, Amit Sharma, Rohan Shah, Matthew Connelly, and Duncan J. Watts. 'Predicting history'. *Nature Human Behaviour* 3.9 (2019): 906–912.

Roodman, David. 'On the probability distribution of long-term changes in the growth rate of the global economy: An outside view.' Working paper, 2020.

Ross, Jacob. 'Rejecting Ethical Deflationism'. *Ethics* 116, no. 4 (2006): 742–768.

Rotemberg, Julio J. 'The Benevolence of the Baker: Fair Pricing under Threat of Customer Anger'. Working paper, 2003.

Rust, John. 'Optimal replacement of GMC bus engines: An empirical model of Harold Zurcher'. *Econometrica* 55, no. 5 (1987): 999–1,033.

Satterthwaite, Mark Allen. 'Strategy-Proofness and Arrow's Conditions: Existence and Correspondence Theorems for Voting Procedures and Social Welfare Functions'. *Journal of Economic Theory* 10, no. 2 (1975): 187–217.

Schall, Lawrence D. 'Interdependent Utilities and Pareto Optimality'. *Quarterly Journal of Economics* 86, no. 1 (1972): 19–24.

Sen, Amartya. *The Idea of Justice*. Harvard, MA: Harvard University Press, 2009.

Sharpe, William F. 'Capital asset prices: A theory of market equilibrium under conditions of risk'. *Journal of Finance* 19, no. 3 (1964): 425–442.

Smith, Nicholas JJ. 'Is evaluative compositionality a requirement of rationality?' *Mind* 123, no. 490 (2014): 457–502.

Snowden, James. 'Should We Give to More Than One Charity?' In *Effective Altruism: Philosophical Issues*, edited by Hilary Greaves and Theron Pummer. Oxford: Oxford University Press, 2019.

Sobel, Joel. 'Interdependent Preferences and Reciprocity'. *Journal of Economic Literature* 43, no. 2 (2005): 392–436.

Stern, Nicholas. *The Economics of Climate Change*. Cambridge, UK: Cambridge University Press, 2006.

Stevenson, Betsey, and Justin Wolfers. 'Economic growth and subjective well-being: Reassessing the Easterlin paradox'. *Brookings Papers on Economic Activity* 39, no. 1 (2008): 1–102.

Stiglitz, Joseph E. *The Price of Inequality: How Today's Divided Society Endangers our Future.* London: Penguin Books, 2013.

Summers, Lawrence and Richard Zeckhauser. 'Policymaking for posterity'. *Journal of Risk and Uncertainty* 37 (2008): 115–140.

Tarsney, Christian. 'Does A Discount Rate Measure The Costs Of Climate Change?' *Economics & Philosophy* 33, no. 3 (2017): 337–365.

———. 'Exceeding Expectations: Stochastic Dominance as a General Decision Theory'. Working paper, 2019.

———. 'The Epistemic Challenge to Longtermism'. Working paper, 2019.

Tarsney, Christian and Teruji Thomas. 'Non-Additive Axiologies in Large Worlds'. Working paper, 2020.

Tännsjö, Torbjörn. 'Future People, the All Affected Principle, and the Limits of the Aggregation Model of Democracy'. In *Homage a Wlodek: Philosophical Papers Dedicated to Wlodek Rabinowicz*, edited by Toni Rønnow-Rasmussen, Björn Petersson, Jonas Josefsson, and Dan Egonsson. Lund, 2007.

Tengs, Tammy O., Miriam E. Adams, Joseph S. Pliskin, Dana Gelb Safran, Joanna E. Siegel, Milton C. Weinstein and John D. Graham. 'Five-Hundred Life-Saving Interventions and Their Cost-Effectiveness'. *Risk Analysis* 15, no. 3 (1994): 369–390.

Tetlock, Philip E. *Expert Political Judgment: How Good Is It? How Can We Know?-New Edition*. Princeton University Press, 2017.

Tetlock, Philip E. and Dan Gardner. *Superforecasting: The Art and Science of Prediction*. Random House, 2016.

Thomas, Teruji. 'The Asymmetry, Uncertainty, and the Long-Term'. Working paper, 2019.

Toit, Jessica du, and Franklin Miller. 'The Ethics of Continued Life-Sustaining Treatment for Those Diagnosed as Brain-Dead'. *Bioethics* 30, no. 3 (2016): 151–158.

Trammell, Philip. 'Patience and Philanthropy'. Manuscript in preparation.

Trammell, Philip and Anton Korinek. 'Economic Growth Under Transformative AI'. Working paper, 2020.

Tran, Brigitte Roth. 'Divest, Disregard, or Double Down?' *American Economic Review*: Insights 1, no. 2 (2019): 241–256.

Vallentyne, Peter and Shelly Kagan. 'Infinite Value and Finitely Additive Value Theory'. *The Journal of Philosophy* 94, no. 1 (1997): 5–26.

Vallinder, Aron and Erik J. Olsson. 'Do computer simulations support the Argument from Disagreement?' *Synthese* 190.8 (2013): 1,437–1,454.

Vivalt, Eva. 'Heterogeneous Treatment Effects in Impact Evaluation'. *American Economic Review, Papers and Proceedings* 105, no. 5 (2015): 467–470.

———. 'How Much Can We Generalize from Impact Evaluations?' *Journal of the European Economic Association*, 2020.

Weitzman, Martin L. 'Why the Far Distant Future Should Be Discounted at Its Lowest Possible Rate'. *Journal of Environmental Economics and Management* 36, no. 3 (1998): 201–208.

———. 'On Modeling and Interpreting the Economics of Catastrophic Climate Change'. *Review of Economics and Statistics* 91, no. 1 (2009): 1–19.

Wilson, Alastair. 'Disagreement, Equal Weight, and Commutativity'. *Philosophical Studies* 149 (2010): 321–326.

Zame, William R. 'Can intergenerational equity be operationalized?' *Theoretical Economics* 2, no. 2 (2007): 187–202.

Appendix A. Research areas for future engagement

This appendix indicates additional areas of possible research that would further GPI's mission, but that GPI itself is not working on now or for the immediately foreseeable future, for reasons of capacity and focus.

A.1. Animal welfare

Given the vast numbers of animals (both wild and farmed) that exist, together with the fact that many animals live in conditions far worse than those faced by the typical human, it is natural to suspect that promoting animal welfare may be among the most cost-effective ways of doing good. Assessing this idea raises a number of interesting and unresolved theoretical questions, including about the ways in which we can improve the world vis-a-vis animal welfare and how we ought to prioritise between interventions that improve human lives and interventions that improve non-human animal lives. These questions are currently particularly neglected within academia.

Potential research projects:

- What sorts of entities have the capacity for sentience? Humans, presumably. But what about non-human animals? Insects? The natural environment? Some forms of artificial intelligence? And how can we make welfare comparisons across them? (This questioning of course takes us beyond specifically animal welfare issues, but the case of non-human animals is a natural place to start this investigation.) How should we make interspecies comparisons of welfare? Is brain size a reasonable proxy? If not, how can we do better?
  PHIL - PHILOSOPHY OF MIND, NEUROETHICS   OTHER - WELFARE BIOLOGY

- Where is the 'zero level' for wellbeing? Which farm animals have lives that are net positive vs net negative? On balance, do wild animals have lives that are net positive or net negative? What are the implications of different population axiologies for this question?
  PHIL - POPULATION ETHICS   OTHER - WELFARE BIOLOGY

- There are consequence-based reasons to promote the consumption of farm animals with higher welfare over those with lower welfare, because such consumption spurs the creation of animals in similar conditions. For hunted meat (most notably wild fish), on the other hand, consumption simply shortens animal lives and/or reduces populations, so it may be better to eat hunted meat of lower welfare. More generally, what are the ethical implications of farmed vs hunted meat consumption?
  PHIL - POPULATION ETHICS   ECON - AGRICULTURAL ECONOMICS, ECOLOGICAL ECONOMICS   OTHER - WELFARE BIOLOGY

- Economic models typically represent animal welfare, if at all, only to the extent that it is represented in human preferences. Can we develop a rigorous economic model that embraces anti-speciesism, and work through how much difference this makes to

the important conclusions such models are used to support, for example within agricultural economics?

ECON - AGRICULTURAL ECONOMICS, WELFARE ECONOMICS

- To what extent, and on what scales, do various shocks to supply and demand (e.g., increased levels of vegetarianism/veganism, bans on battery cages) affect the number of animals farmed for food (in total and/or under given welfare conditions)?

ECON - AGRICULTURAL ECONOMICS, INDUSTRIAL ORGANISATION

- To what extent would changes to the farm production of one animal affect the numbers of other (farmed and wild) animals born?

ECON - AGRICULTURAL ECONOMICS, INDUSTRIAL ORGANISATION   OTHER - ECOLOGY

- Consider the 'meat eater' problem: interventions that save human lives and/or boost economic growth have obvious direct benefits, but both lead to increases in the consumption and production of animal products. To what extent could this significantly decrease the net positive impact of such interventions, or even imply that they have net negative value? To what extent do positive indirect effects of economic growth push in the other direction? Which of these sets of considerations is larger, when all indirect effects are counted? How, in general, should we think about the impact on animals of improving human lives?

ECON - AGRICULTURAL ECONOMICS, GROWTH

- What is the case for the claim that improving the living conditions of non-human animals in the wild is among the most cost-effective causes? What tractable activities are there aimed at promoting such improvements?

ECON - ECOLOGICAL ECONOMICS   OTHER - WELFARE BIOLOGY

Existing academic literature:

- Cowen, Tyler. 'Policing Nature'. *Environmental Ethics* 25, no. 2 (2003): 169–182.

- Delon, Nicolas, and Duncan Purves. 'Wild Animal Suffering is Intractable'. *Journal of Agricultural and Environmental Ethics* 31, no. 2 (2018): 239–260.

- Lusk, J. L., and F. B. Norwood. 'Animal Welfare Economics'. *Applied Economic Perspectives and Policy* 33, no. 4 (2011): 463–483.

- ———. 'Speciesism, Altruism and the Economics of Animal Welfare'. *European Review of Agricultural Economics* 39, no. 2 (2012): 189–212.

- Malone, Trey and Jayson Lusk. 'Putting the Chicken Before the Egg Price: An Ex Post Analysis of California's Battery Cage Ban'. *Journal of Agricultural & Resource Economics* 41, no. 3 (2016): 518–532.

- Matheny, Gaverick. 'Least Harm: A Defense of Vegetarianism from Steven Davis's Omnivorous Proposal'. *Journal of Agricultural and Environmental Ethics* 16, no. 5 (2003): 505–511.

- Matheny, Gaverick and Kai M. A. Chan. 'Human Diets and Animal Welfare: The Illogic of the Larder'. *Journal of Agricultural and Environmental Ethics* 18, no. 6 (2005): 579–594.

- McMahan, Jeff. 'The Moral Problem of Predation'. In *Philosophy Comes to Dinner: Arguments About the Ethics of Eating*, edited by Andrew Chignell, Terence Cuneo, and Matthew C. Halteman. New York: Routledge, 2015. 268–294.

- Ng, Yew-Kwang. 'Towards Welfare Biology: Evolutionary Economics of Animal Consciousness and Suffering'. *Biology & Philosophy* 10, no. 3 (1995): 255–285.

- Norwood, F. Bailey and Jayson Lusk. *Compassion, by the Pound: The Economics of Farm Animal Welfare*. New York: Oxford University Press, 2011.

- Singer, Peter. 'Speciesism and Moral Status'. *Metaphilosophy* 40, no. 3–4 (2009): 567–581.

- Višak, Tatjana and Robert Garner, eds. *The Ethics of Killing Animals*. New York: Oxford University Press, 2015.

Existing informal discussion:

- Rossa O'Keeffe-O'Donovan and Eva Vivalt, Animal Agriculture and Economics, 2017.

- Carl Shulman, How are brain mass (and neurons) distributed among humans and the major farmed land animals?, 2013.

- Carl Shulman, Trends in farmed animal life-years per kg and per human in the United States, 2015.

- Carl Shulman, Various functional forms for brain-weighting wild insects and farmed land animals favor the former, 2015.

- Carl Shulman, Vegan advocacy and pessimism about wild animal welfare, 2013.

- Carl Shulman, Some considerations for prioritization within animal agriculture, 2015.

- Brian Tomasik, Is Brain Size Morally Relevant?, 2020.

- Luke Muehlhauser, 2017 Report on Consciousness and Moral Patienthood, 2017.

- Brian Tomasik, The Importance of Wild Animal Suffering, 2019.

- Michael Dickens, Why the Open Philanthropy Project Should Prioritize Wild Animal Suffering, 2016.

- Toby Ord, Crucial Considerations for Animal Welfare, 2017.

- Robin Hanson, Why Meat is Moral, and Veggies are Immoral, 2002.

A.2. The scope of welfare maximisation

This topic concerns whether impartial welfare maximisation is simply a beneficial project that one might or might not choose to engage in, or whether stronger things can be said in its favour from the point of view of moral philosophy.

Potential research projects:

- If it is the case that the long-run effects of one's actions are much larger in impact than the short-run effects, does this strengthen the case for there being strong duties of beneficence, simply because altruistic actions do so much more good than we might have thought?

PHIL - DUTIES OF BENEFICENCE

- Non-consequentialist views often make 'emergency situation' provisos, where they tend to make recommendations in a more consequentialist manner (such as permitting rights violations or making acts of altruism obligatory). To what extent is it justified to think that we are living in an 'emergency situation'?

PHIL - DUTIES OF BENEFICENCE, DEONTOLOGICAL CONSTRAINTS

- If there is an obligation to engage in impartial welfare maximisation, what is the nature of that obligation? Should all our resources be spent in whichever way would do the most good? For example, is it that there is an obligation to maximise the effectiveness of whatever sacrifices one makes, but (at least beyond a certain point) no obligation to make the sacrifices?

PHIL - DUTIES OF BENEFICENCE, CONDITIONAL OBLIGATION

- Do obligations of beneficence require cause-neutrality?

PHIL - DUTIES OF BENEFICENCE

- Even if beneficence is only one of many competing obligations in our lives, is it still the case that with respect to the reasons of beneficence that we have, we ought to try to do the most good?

PHIL - DUTIES OF BENEFICENCE

- It is often claimed that all plausible moral theories recognise a pro tanto reason to promote the impartial good. To what extent does this claim justify the further claim that the project of impartial benevolence, and the associated research questions (as described in this research agenda), are important by the lights of all plausible moral theories?

PHIL - MORAL UNCERTAINTY, DUTIES OF BENEFICENCE

Existing academic literature:

- Horton, Joe. 'The All or Nothing Problem'. *The Journal of Philosophy* 114, no. 2 (2017): 94–104.

- Mogensen, Andreas L. 'Should We Prevent Optimific Wrongs?' *Utilitas* 28, no. 2 (2016): 215–226.

- Pummer, Theron. 'Whether and Where to Give'. *Philosophy & Public Affairs* 44, no. 1 (2016): 77–95.

- Singer, Peter. *The Life You Can Save: Acting Now to End World Poverty.* New York: Random House, 2009.

- ———. *The Most Good You Can Do: How Effective Altruism Is Changing Ideas about Living Ethically.* New Haven: Yale University Press, 2015.

Existing informal discussion:

- Theron Pummer, [People and charitable causes are importantly different things](), 2014.

Appendix B. Closely related areas of existing academic research

Here we indicate areas of existing academic literature that serve as particularly relevant background for the topics on this research agenda. Interested researchers who also have background expertise in one or more of these areas are likely to be particularly good fits to GPI's research agenda.

B.1. Methodology of cost-benefit analysis and cost-effectiveness analysis

Cost-benefit analysis (CBA) and cost-effectiveness analysis (CEA) are standard tools for evaluating projects. Several aspects of the methodology of CBA and CEA, however, are contested, often for reasons that tap into fundamental normative controversies. Examples include the choice of a pure time discount rate in trading off costs/benefits incurred earlier against those incurred later, and the use or not of 'distributional weights' (e.g., to account for the fact that a marginal dollar is worth more to a poor person than to a rich person).

Examples of relevant literature:

- Adler, Matthew D. 'Benefit-Cost Analysis and Distributional Weights: An Overview'. *Review of Environmental Economics and Policy* 10, no. 2 (2016): 264–285.

- Bronsteen, John, Christopher Buccafusco and Jonathan Masur. 'Well-Being Analysis vs. Cost-Benefit Analysis'. *Duke Law Journal* 62 (2013): 1,603–1,689.

- Sen, Amartya. 'The Discipline of Cost-Benefit Analysis'. *The Journal of Legal Studies* 29, no. S2 (2000): 931–952.

- J-PAL. Conducting Cost-Effectiveness Analysis.

- HM Treasury. *The Green Book: Appraisal and Evaluation in Central Government*. London: TSO, 2013.

B.2. Multidimensional economic indices

A number of efforts have been made in the last decade or so to come up with macroeconomic measures that capture more than GDP. Some, for example, incorporate 'environmental capital', or value biodiversity loss, in addition to accounting for the resources already under human ownership and in productive use. Relatedly, a literature in development economics focuses on constructing 'multidimensional poverty indices', which define poverty in terms not only of income or consumption, but also other factors for which income may serve as an incomplete proxy: factors such as years of schooling, quality of housing, longevity or literacy. In general, multidimensional indices are useful for accounting for the full impacts of any set of interventions, but they are particularly important to the project of comparing interventions across very different causes.

Examples of relevant literature:

- Indices involving environmental capital, etc.:
  - Daly, Herman E., John B. Cobb Jr and John B. Cobb. *For the Common Good: Redirecting the Economy toward Community, the Environment, and a Sustainable Future*. Boston: Beacon Press, 1994.
  - Anielski, Mark. 'Measuring the Sustainability of Nations: The Genuine Progress Indicator System of Sustainable Well Being Accounts'. *The Fourth Biennial Conference of the Canadian Society for Ecological Economics: Ecological Sustainability of the Global Marketplace*. 2001.
- The Multidimensional Poverty Index:
  - Alkire, Sabina and James Foster. 'Counting and Multidimensional Poverty Measurement'. *Journal of Public Economics* 95, no. 7 (2011): 476–487.
- Examples of prior approaches to generating multidimensional poverty indices:
  - Axiomatic:
    - Chakravarty, Satya R., Diganta Mukherjee and Ravindra R. Ranade. 'On the Family of Subgroup and Factor Decomposable Measures of Multidimensional Poverty'. *Research on Economic Inequality* 8 (1998): 175–194.
  - Information-theoretic:
    - Maasoumi, Esfandiar and Maria Ana Lugo. 'The Information Basis of Multivariate Poverty Assessments'. In *Quantitative Approaches to Multidimensional Poverty Measurement*, edited by Nanak Kakwani and Jacques Silber. London: Palgrave Macmillan, 2008. 1–29.
  - Fuzzy set:
    - Cerioli, Andrea and Sergio Zani. 'A Fuzzy Approach to the Measurement of Poverty'. In *Income and Wealth Distribution, Inequality and Poverty*, edited by Camilo Dagum and Michele Zenga. Berlin; Heidelberg: Springer-Verlag, 1990. 272–284.
  - Latent variable:
    - Kakwani, Nanak and Jacques Silber, eds. *Quantitative Approaches to Multidimensional Poverty Measurement*. London: Palgrave Macmillan, 2008.
- 'Capability approach':
  - Nussbaum, Martha C. *Creating Capabilities*. Cambridge, MA: Harvard University Press, 2011.
  - Sen, Amartya Kumar. *Commodities and Capabilities*. Amsterdam: North-Holland, 1985.

B.3. Infinite ethics and intergenerational equity

It is conceivable, and in fact implied by some contemporary cosmological theories, that the universe contains an infinite number of potentially value-bearing entities, such as happy and sad people, and therefore an infinite amount of positive and/or negative value. If no action can affect more than a finite amount of value, it follows in standard cardinal arithmetic that no action can affect the value of the world. This raises the question of how such 'infinitarian paralysis' can be avoided. Alternatively, if some of our actions may have consequences of infinite value, and if we do not render them finite by discounting – that is, if we act on some principle of 'intergenerational equity' – we face the question of how to compare such consequences, or probabilities of such consequences.

Examples of relevant literature:

- Arntzenius, Frank. '[Utilitarianism, Decision Theory, and Eternity](#)'. *Philosophical Perspectives* 28, no. 1 (2014): 31–58.

- Asheim, Geir B. '[Intergenerational Equity](#)'. *Annual Review of Economics* 2, no. 1 (2010): 197–222.

- Basu, Kaushik and Tapan Mitra. '[Aggregating Infinite Utility Streams with Intergenerational Equity: The Impossibility of Being Paretian](#)'. *Econometrica* 71, no. 5 (2003): 1,557–1,563.

- Bostrom, Nick. '[Infinite Ethics](#)'. *Analysis and Metaphysics*, no. 10 (2011): 9–59.

- Lauwers, Luc and Peter Vallentyne. '[Infinite Utilitarianism: More Is Always Better](#)'. *Economics & Philosophy* 20, no. 2 (2004): 307–330.

- Vallentyne, Peter and Shelly Kagan. '[Infinite Value and Finitely Additive Value Theory](#)'. *The Journal of Philosophy* 94, no. 1 (1997): 5–26.

- Zame, William R. '[Can intergenerational equity be operationalized?](#)' *Theoretical Economics* 2, no. 2 (2007): 187–202.

B.4. Epistemology of disagreement

Given our state of uncertainty, many topics within global priorities research will inevitably be subject to disagreement among intelligent and well-informed people. As a result, we must often deal with the question of how to act in the face of disagreement among 'epistemic peers': those of roughly equal competence with respect to the question at hand. This question has been studied extensively both in the abstract and with explicit reference to contentious issues central to global prioritisation, such as the social discount rate.

Examples of relevant literature:

- Christensen, David. '[Epistemology of Disagreement: The Good News](#)'. *Philosophical Review* 116, no. 2 (2007): 187–217.

- Christensen, David. '[Disagreement as Evidence: The Epistemology of Controversy](#)'. *Philosophy Compass* 4, no. 5 (2009): 756–767.

- Christensen, David and Jennifer Lackey, eds. *The Epistemology of Disagreement: New Essays*. Oxford: Oxford University Press, 2013.

- Elga, Adam. '[Reflection and Disagreement](#)'. *Noûs* 41, no. 3 (2007): 478–502.

- Feldman, Richard and Ted A. Warfield, eds. *Disagreement*. Oxford: Oxford University Press, 2010.

- Freeman, Mark C. and Ben Groom. '[Positively Gamma Discounting: Combining the Opinions of Experts on the Social Discount Rate](#)'. *The Economic Journal* 125, no. 585 (2015): 1,015–1,024.

- Jouini, Elyes, Jean-Michel Marin and Clotilde Napp. '[Discounting and Divergence of Opinion](#)'. *Journal of Economic Theory* 145, no. 2 (2010): 830–859.

- Wilson, Alastair. '[Disagreement, Equal Weight, and Commutativity](#)'. *Philosophical Studies* 149 (2010): 321–326.

B.5. Demandingness

Maximising consequentialism is sometimes objected to on the grounds that it is overly demanding. For example, going out for dinner at a mid-range restaurant is seen as a permissible option by 'common-sense morality', but such an action is unlikely to have the best consequences impartially considered, and is therefore judged impermissible by maximising consequentialism. Research into the scope of individuals' and institutions' moral obligations toward global welfare maximisation must therefore contend with such demandingness objections.

Examples of relevant literature:

- Scheffler, Samuel. *The Rejection of Consequentialism*. Oxford: Clarendon Press, 1982.

- Kagan, Shelly. *The Limits of Morality*. Oxford: Clarendon Press, 1989.

- Unger, Peter. *Living High and Letting Die*. New York: Oxford University Press, 1996.

B.6. Population ethics

Our relative evaluations of projects across many cause areas depend to a large extent on our understanding of how to compare outcomes in which different groups of individuals may exist. Answers to questions in population ethics appear particularly important regarding questions about the value of extinction risk reduction, about the value of farm animal welfare efforts, and about whether to save or improve lives.

Examples of relevant literature:

- Arrhenius, Gustaf. 'Population Ethics: The Challenge of Future Generations'. Manuscript in preparation.

- Greaves, Hilary. 'Population axiology'. *Philosophy Compass* 12, no. 11 (2017): e12442.
- Parfit, Derek. *Reasons and Persons*. Oxford: Oxford University Press, 1984.

B.7. Risk aversion and ambiguity aversion

Our uncertainty about activities' long-term consequences can differ widely by cause area. Risk aversion can therefore substantially affect the decision of whether, for example, to prioritise reductions in existential risk or in near-term suffering. Because the *precision* of our beliefs about long-term consequences can also differ widely, ambiguity aversion can affect our prioritisation decisions similarly. The question of global prioritisation therefore relies heavily on the question of whether, and to what extent, we ought to avoid risk and ambiguity.

Examples of relevant literature:

- On risk aversion:
  - Surveys with some empirical emphasis:
    - Chandler, Jake. 'Descriptive Decision Theory'. *The Stanford Encyclopaedia of Philosophy (Winter 2017 Edition). Metaphysics Research Lab, Stanford University*, 2017.
    - Fox, Craig R., Carsten Erner and Daniel J. Walters. 'Decision Under Risk: From the Field to the Laboratory and Back'. In *The Wiley Blackwell Handbook of Judgment and Decision Making*, edited by Gideon Keren and George Wu. Chichester, UK: John Wiley & Sons, 2015: 41–88.
  - Recent normative theories and discussions:
    - Buchak, Lara. *Risk and Rationality*. Oxford: Oxford University Press, 2013.
    - Stefánsson, H. Orri and Richard Bradley. 'What is Risk Aversion?' *British Journal for the Philosophy of Science*, forthcoming.
  - Other discussions related to risk aversion in expected utility theory:
    - Allais, Maurice. 'The Foundations of a Positive Theory of Choice Involving Risk and a Criticism of the Postulates and Axioms of the American School'. *Econometrica* 21, no. 4 (1953): 503–546.
    - Hansson, Bengt. 'Risk aversion as a Problem of Conjoint Measurement'. In *Decision, Probability and Utility: Selected Readings*, edited by Peter Gärdenfors and Nils-Eric Sahlin. Cambridge, UK: Cambridge University Press, 1988. 136–158.
    - Rabin, Matthew. 'Risk Aversion and Expected-Utility Theory: A Calibration Theorem'. *Econometrica* 68, no. 5 (2000): 1,281–1,292.
    - Rothschild, Michael and Joseph E. Stiglitz. 'Increasing Risk: I. A Definition'. *Journal of Economic Theory* 2, no. 3 (1970): 225–243.

- Cumulative prospect theory and related descriptive theories:
  - Wakker, Peter P. *Prospect Theory: For Risk and Ambiguity.* Cambridge, UK: Cambridge University Press, 2010.
- On ambiguity aversion:
  - Ellsberg, Daniel. 'Risk, Ambiguity, and the Savage Axioms'. *Quarterly Journal of Economics* 75, no. 4 (1961): 643–669.
  - Models of ambiguity aversion:
    - Etner, Johanna, Meglena Jeleva and Jean-Marc Tallon. 'Decision Theory under Ambiguity'. *Journal of Economic Surveys* 26, no. 2 (2012): 234–270.
  - Normative discussion of ambiguity aversion:
    - Al-Najjar, Nabil I. and Jonathan Weinstein. 'The Ambiguity Aversion Literature: A Critical Assessment'. *Economics & Philosophy* 25, no. 3 (2009): 249–284.
    - Bradley, Richard. *Decision Theory with a Human Face.* Cambridge, UK: Cambridge University Press, 2017.
    - Raiffa, Howard. 'Risk, Ambiguity, and the Savage Axioms: Comment'. *Quarterly Journal of Economics* 75, no. 4 (1961): 690–694.
    - Siniscalchi, Marciano. 'Two Out Of Three Ain't Bad: A comment on "the ambiguity aversion literature: A critical assessment"'. *Economics & Philosophy* 25, no. 3 (2009): 335–356.
    - Row, Thomas and Alex Voorhoevee. 'Egalitarianism under Severe Uncertainty'. *Philosophy & Public Affairs* 36, no.3 (2018): 239–268.
  - Experimental literature (descriptive rather than normative):
    - Trautmann, Stefan and Gijs van de Kuilen. 'Ambiguity Attitudes'. In *The Wiley-Blackwell Handbook of Judgment and Decision Making*, edited by Gideon Keren and George Wu. Chichester, UK: John Wiley & Sons, 2015: 89–116.

B.8. Moral uncertainty

Attempts to compare the importance of different problems or the effectiveness of different interventions, for example, in programme evaluation research in economics, often default to using a utilitarian framework. But, even if one is sympathetic to utilitarianism, it would clearly be overconfident to be *certain* in that moral theory. So, plausibly, we should try to incorporate moral uncertainty into our reasoning when we prioritise among problems. This raises the general question of what form appropriate action under moral uncertainty takes. A framework for action under moral uncertainty is ultimately necessary for resolving questions regarding which causes are most important, given said moral uncertainty; regarding whether and in what way it is permissible to cause harm in the course of doing good; and regarding the extent of individuals' and institutions' obligations toward impartial benevolence (including, e.g., benevolence toward individuals in the distant future).

Examples of relevant literature:

- Greaves, Hilary and Toby Ord. 'Moral Uncertainty About Population Axiology'. *Journal of Ethics & Social Philosophy* 12, no. 2 (2017): 135–167.

- Gustafsson, Johan E. and Olle Torpman. 'In Defence of My Favourite Theory'. *Pacific Philosophical Quarterly* 95, no. 2 (2014): 159–174.

- Lockhart, Ted. *Moral Uncertainty and Its Consequences*. Oxford: Oxford University Press, 2000.

- Harman, Elizabeth. 'Does Moral Ignorance Exculpate?' *Ratio* 24, no. 4 (2011): 443–468.

- MacAskill, William, Krister Bykvist and Toby Ord. *Moral Uncertainty*. Oxford: Oxford University Press, 2020.

- MacAskill, William. 'The Infectiousness of Nihilism'. *Ethics* 123, no. 3 (2013): 508–520.

- MacAskill, William, Owen Cotton-Barratt and Toby Ord. 'Statistical normalization methods in interpersonal and intertheoretic comparison'. *Journal of Philosophy* 117 no. 2 (2020): 61–95.

- Mason, Elinor. 'Moral Ignorance and Blameworthiness'. *Philosophical Studies* 172, no. 11 (2015): 3,037–3,057.

- Ross, Jacob. 'Rejecting Ethical Deflationism'. *Ethics* 116, no. 4 (2006): 742–768.

- Sepielli, Andrew. 'Along an Imperfectly-Lighted Path'. PhD dissertation. New Brunswick: Rutgers University, 2010.

- ———. 'Moral Uncertainty and the Principle of Equity among Moral Theories'. *Philosophy and Phenomenological Research* 86, no. 3 (2013): 580–589.

- Weatherson, Brian. 'Running Risks Morally'. *Philosophical Studies* 167, no. 1 (2014): 141–163.

B.9. Value of information

In situations of uncertainty, information can greatly increase our chances of choosing better actions. The timelines on which we expect to acquire information, the costs of acquiring it and the extent to which we expect that it will be action-guiding can all affect our decisions concerning, for example, whether to commit resources sooner or later. More generally, considerations regarding the value of information inform the importance we place on the 'option value' of delaying any irreversible development of unknown value, such as human extinction, until after more information has been acquired.

Examples of relevant literature:

- Arrow, Kenneth J. and Anthony C. Fisher. 'Environmental Preservation, Uncertainty, and Irreversibility'. *Quarterly Journal of Economics* 88, no. 2 (1974): 312–319.

- Bishop, Richard C. 'Option Value: An Exposition and Extension'. *Land Economics* 58, no. 1 (1982): 1–15.

- Dixit, Avinash and Robert S. Pindyck. *Investment Under Uncertainty*. Princeton: Princeton University Press, 1994.

- Eeckhoudt, Louis and Philippe Godfroid. 'Risk Aversion and the Value of Information'. *The Journal of Economic Education* 31, no. 4 (2000): 382–388.

- Good, Irving John. 'On the Principle of Total Evidence'. *The British Journal for the Philosophy of Science* 17, no. 4 (1967): 319–321.

B.10. Harnessing and combining evidence

When choosing among approaches to working on a particular problem or cause area, individuals and organisations should use empirical evidence to estimate which approach will be most effective. In some fields, for example, in development economics, there has been a large increase in the availability of high-quality studies, including randomised controlled trials estimating the effect of different interventions or programmes. However, it is often not clear how to combine information from different studies, particularly when they were undertaken in different settings or use different empirical methods, even if they are evaluating essentially the same intervention. For other questions of interest, it is inherently more difficult (and sometimes impossible) to run randomised trials and we must use information from other sources, including theoretical models and other types of empirical evidence, to make informed judgements. General research into how best to harness and combine the available sources of evidence therefore has broad relevance to the enterprise of global prioritisation.

Examples of relevant literature:

- Deaton, Angus and Nancy Cartwright. 'Understanding and Misunderstanding Randomized Controlled Trials'. *Social Science & Medicine* 210 (2018): 2–21.

- External validity:

    - Bold, Tessa et al. 'Scaling Up What Works: Experimental Evidence on External Validity in Kenyan Education'. *Center for Global Development* Working Paper 321, 2013.

    - Dehejia, Rajeev, Cristian Pop-Eleches and Cyrus Samii. 'From Local to Global: External Validity in a Fertility Natural Experiment'. *IZA Discussion Papers* 9300, 2015.

    - Vivalt, Eva. 'How Much Can We Generalize from Impact Evaluations?' *Journal of the European Economic Association*, 2020.

- New approaches to drawing inferences out of sample:

    - Chassang, Sylvain, Padró I. Miquel and Erik Snowberg. 'Selective Trials: A Principal-Agent Approach to Randomized Controlled Experiments'. *American Economic Review* 102, no. 4 (2012): 1,279–1,309.

- Gechter, Michael. '[Generalizing the Results from Social Experiments: Theory and Evidence from Mexico and India](#)'. Manuscript in preparation.

- Gechter, Michael, Cyrus Samii, Rajeev Dehejia and Cristian Pop-Eleches. '[Evaluating Ex Ante Counterfactual Predictions Using Ex Post Causal Inference](#)'. Working paper, 2019.

- Kowalski, Todd J., Shanu N. Kothari, Michelle A. Mathiason and Andrew J. Borgert. '[Impact of Hair Removal on Surgical Site Infection Rates: A Prospective Randomized Noninferiority Trial](#)'. *Journal of the American College of Surgeons* 223, no. 5 (2016): 704–711.

- Structural modelling:

  - Attanasio, Orazio P., Costas Meghir, and Ana Santiago. '[Education Choices in Mexico: Using a Structural Model and a Randomized Experiment to Evaluate Progress](#)'. *The Review of Economic Studies* 79, no. 1 (2011): 37–66.

  - Keane, Michael P. '[Structural vs. Atheoretic Approaches to Econometrics](#)'. *Journal of Econometrics* 156, no. 1 (2010): 3–20.

  - Low, Hamish and Costas Meghir. '[The Use of Structural Models in Econometrics](#)'. *Journal of Economic Perspectives* 31, no. 2 (2017): 33–58.

- Qualitative evidence:

  - Bennett, Andrew and Jeffrey T. Checkel. '[Process Tracing: From Philosophical Roots to Best Practices](#)'. In *Process Tracing: From Metaphor to Analytic Tool*, edited by Andrew Bennett and Jeffrey T. Checkel. Cambridge, UK: Cambridge University Press, 2014. 3–37.

  - Freedman, David A. '[On Types of Scientific Enquiry: The Role of Qualitative Reasoning](#)'. In *Oxford Handbook of Political Methodology*, edited by Janet M. Box-Steffensmeier, Henry E. Brady and David Collier. Oxford: Oxford University Press, 2008. 300–318.

  - Goertz, Gary and James Mahoney. *A Tale of Two Cultures: Qualitative and Quantitative Research in the Social Sciences*. Princeton: Princeton University Press, 2012.

  - Seawright, Jason. *Multi-Method Social Science: Combining Qualitative and Quantitative Tools*. Cambridge, UK: Cambridge University Press, 2016.

B.11. The psychology of altruistic decision-making

Various apparently altruistic and reasonable behaviours seem puzzling on closer inspection, if we assume that the agent is attempting to maximise the expected impact of their actions. These behaviours include (a) donating to more than one charity and (b) avoiding supporting work on mitigating existential risks on the grounds of 'risk aversion'. The same behaviours might make more sense assuming a less pure form of altruism (the most obvious alternative being a 'warm glow' theory of motivation), or assuming deviations from expected utility theory that are arguably irrational (such as ambiguity aversion and certain forms of risk aversion).

A better understanding of the variety of psychological mechanisms underlying altruistic behaviour might aid efforts to work around behavioural limitations, and maximise the good done by imperfectly altruistic agents.

Examples of relevant literature:

- Andreoni, James. 'Privately Provided Public Goods in a Large Economy: The Limits of Altruism'. *Journal of Public Economics* 35 (1988): 57–73.

- ———. 'Giving with Impure Altruism: Applications to Charity and Ricardian Equivalence'. *Journal of Political Economy* 97 (1989): 1,447–1,458.

- ———. 'Impure Altruism and Donations to Public Goods: A Theory of Warm-Glow Giving'. *The Economic Journal* 100 (1990): 464–477.

- Ashraf, Nava, and Oriana Bandiera. 'Altruistic Capital'. *American Economic Review* 107, no. 5 (2017): 70–75.

- Batson, Charles Daniel. *Altruism in Humans*. New York: Oxford University Press, 2011.

- Batson, C. Daniel, Nadia Ahmad, David A. Lishner and Jo-Ann Tsang. 'Empathy and Altruism'. In *Oxford Handbook of Positive Psychology*, edited by C. R. Snyder and Shane J. Lopez. Oxford: Oxford University Press, 2009. 417–427.

- Bloom, Paul. *Against Empathy: The Case for Rational Compassion*. New York: Harper Collins, 2016.

- Caviola, Lucius, Nadira Faulmüller, Jim AC Everett, Julian Savulescu and Guy Kahane. 'The Evaluability Bias in Charitable Giving: Saving Administration Costs or Saving Lives?' *Judgment and Decision Making* 9, no. 4 (2014): 303–315.

- Elster, Jon. 'The Valmont Effect: The Warm-Glow Theory of Philanthropy'. In *Giving Well: The Ethics of Philanthropy*, edited by Patricia M. L. Illingworth, Thomas Pogge and Leif Wenar. New York: Oxford University Press, 2011. 67–83.

- Fong, Christina M. and Felix Oberholzer-Gee. 'Truth in Giving: Experimental Evidence on the Welfare Effects of Informed Giving to the Poor'. *Journal of Public Economics* 95, no. 5–6 (2011): 436–444.

- Karlan, Dean, John A. List and Eldar Shafir. 'Small Matches and Charitable Giving: Evidence from a Natural Field Experiment'. *Journal of Public Economics* 95, no. 5–6 (2011): 344–350.

- Null, Clair. 'Warm Glow, Information, and Inefficient Charitable Giving'. *Journal of Public Economics* 95, no. 5–6 (2011): 455–465.

- Oakley, Barbara, Ariel Knafo, Guruprasad Madhavan and David Sloan Wilson, eds. *Pathological Altruism*. Oxford: Oxford University Press, 2011.

- Povey, Richard. 'The Limits to Altruism - A Survey'. Working paper, 2014.

- ———. 'The Socially Optimal Level of Altruism'. Working paper, 2015.

- Simler, Kevin, and Robin Hanson. *The Elephant in the Brain: Hidden Motives in Everyday Life*. Oxford: Oxford University Press, 2017.

- Small, Deborah A., and George Loewenstein. 'Helping *a* Victim or Helping *the* Victim: Altruism and Identifiability'. *Journal of Risk and Uncertainty* 26, no. 1 (2003): 5–16.

- Smith, Sarah, Frank Windmeijer and Edmund Wright. 'Peer Effects in Charitable Giving: Evidence from the (Running) Field'. *The Economic Journal* 125, no. 585 (2014): 1,053–1,071.

Appendix C. Additional informal discussion

This appendix contains links to additional informal discussion of the themes discussed in this research agenda.

A good introductory overview of the theoretical side of global priorities research is [Prospecting for Gold](#) by Owen Cotton-Barratt.

The most important websites to get up to speed on current thought and debates in the effective altruism community are as follows:

- https://www.givewell.org/, and their blog
- https://www.openphilanthropy.org/, and their blog
- https://www.80000hours.org/, and their blog
- http://globalprioritiesproject.org/
- https://concepts.effectivealtruism.org/
- https://www.effectivealtruism.org/articles/
- http://reducing-suffering.org/
- https://longtermrisk.org/
- https://rationalaltruist.com/
- http://reflectivedisequilibrium.blogspot.co.uk/
- https://forum.effectivealtruism.org/ (Though this also contains discussion of effective altruism community issues that aren't as relevant to GPI's research agenda.)
- https://www.lesswrong.com/ (Though this also contains discussion of issues concerning rationality that aren't as relevant to GPI's research agenda.)

Finally, here is an incomplete list of some of the most important articles and blog posts from the effective altruism community that are relevant to GPI's research agenda (many of which are also mentioned above):

- Scott Alexander, [Ethics offsets](#), 2015.
- Scott Alexander, [Nobody is perfect, everything is commensurable](#), 2014.
- David Althaus and Lukas Gloor, [Reducing Risks of Astronomical Suffering: A Neglected Priority](#), 2019.
- Nick Beckstead, [On the overwhelming importance of shaping the far future](#), 2013.
- Nick Beckstead, [A proposed adjustment to the astronomical waste argument](#), 2013.
- Nick Bostrom, [3 ways to advance science](#), 2012.
- Nick Bostrom, [Crucial considerations and wise philanthropy](#), 2014.
- Paul Christiano, [Astronomical waste](#), 2013.

- Paul Christiano, Influencing the far future, 2013.
- Paul Christiano, Neglectedness and impact, 2014.
- Paul Christiano, Pressing ethical questions, 2014.
- Paul Christiano, Replaceability, 2013.
- Paul Christiano, The best reason to give later, 2013.
- Paul Christiano, The efficiency of modern philanthropy, 2013.
- Owen Cotton-Barratt, How valuable is movement growth?, 2015.
- Owen Cotton-Barratt and Ben Todd, Give now or later?, 2015.
- Katja Grace, Cause Prioritization Research, undated.
- Katja Grace, Estimation Is the Best We Have, 2011.
- Robin Hanson, Marginal Charity, 2012.
- Robin Hanson, Parable of the Multiplier Hole, 2010.
- Holden Karnofsky, Flow-through effects, 2013.
- Holden Karnofsky, Hits-Based Giving, 2016.
- Holden Karnofsky, Passive vs. rational vs. quantified, 2013.
- Holden Karnofsky, Sequence thinking vs. cluster thinking, 2016.
- Holden Karnofsky, Why we can't take expected value estimates literally even when they're unbiased, 2016.
- Holden Karnofsky, Worldview diversification, 2016.
- Jeff Kaufman, Altruism isn't about sacrifice, 2013.
- Jeff Kaufman, The Unintuitive Power Laws of Giving, 2013.
- Ben Kuhn, A critique of effective altruism, 2013.
- Greg Lewis, Beware Surprising and Suspicious Convergence, 2016.
- Toby Ord, The Moral Imperative Towards Cost-Effectiveness in Global Health, 2013.
- Carl Shulman, Are pain and pleasure equally energy efficient?, 2012.
- Carl Shulman and Nick Beckstead, A long-run perspective on strategic cause selection and philanthropy, 2013.
- Jonah Sinick, Many Weak Arguments vs. One Relatively Strong Argument, 2013.
- Scott Siskind, Efficient charity: do unto others…, 2013.
- Benjamin Todd, List of the most urgent global issues, 2018.
- Benjamin Todd, The value of coordination, 2016.
- Brian Tomasik, Charity Cost Effectiveness in an Uncertain World, 2015.
- Brian Tomasik, The Haste Consideration, Revisited, 2018.
- Brian Tomasik, Two-Envelopes Problem for Brain Size and Moral Uncertainty, 2018.

- Brian Tomasik, [Why Charities Usually Don't Differ Astronomically in Expected Cost-Effectiveness](#), 2017.

- Brian Tomasik, [How the Simulation Argument Dampens Future Fanaticism](#), 2018.

- James Steinhardt, [Another Critique of Effective Altruism](#), 2014.

- Robert Wiblin, [How to create the world's most effective charity](#), 2013.

- Robert Wiblin, [One approach to comparing global problems in terms of expected impact](#), 2019.

globalprioritiesinstitute.org