A RESEARCH AGENDA FOR THE GLOBAL PRIORITIES INSTITUTE

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.

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.

Contents

Global Priorities Institute (GPI)

Stakeholder(s):

Actors :
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.

Non-Consequentialists :
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.

Effective Altruism Community :
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.

Open Philanthropy Project :
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.

Academics :
However, many of these ideas have yet to be explored in academia. | 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.

Economists

Philosophers

Research Contributors

Hilary Greaves

William MacAskill

Rossa O'Keeffe-O'Donovan

Philip Trammell

Benjamin Tereick

Andreas Mogensen

Christian Tarsney

Gustav Alexandrie

Maxime Cugnon de Sévricourt

Other Contributors

James Aung

Luzia Bruckamp

John Mori

Laura Pomarius

Sven Herrmann

Other Assistants :
We would like to thank Pablo Stafforini, Aron Vallinder, the Global Priorities Institute Advisory Board, and numerous colleagues at the Future of Humanity Institute, the Centre for Effective Altruism and elsewhere for their invaluable assistance in composing this agenda.

Pablo Stafforini

Aron Vallinder

Global Priorities Institute Advisory Board

Future of Humanity Institute Colleagues

Centre for Effective Altruism Colleagues

Vision

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

Mission

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

1. Longtermism

The longtermism paradigm ~ ... 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 & Evaluation

Articulate and evaluate longtermism

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.

1.2. Future of Humanity

Consider the value of the future of humanity

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.

1.3. Catastrophic Risk

Reduce and mitigate catastrophic risk

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.

1.4. Opportunities

Identify trajectory-change opportunities

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.

1.5. Intergenerational Governance

Evaluate the long-term consequences of policies and actions

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.

Stakeholder(s):

Future Generations **Philanthropists**

Policymakers

1.5.1. Influence

Influence the behaviour of future policymakers

Stakeholder(s):

Future Policymakers

1.5.2. Coordination

Coordinate current actions in recognition of the constraints on influencing future policy makers

1.6. Growth & Inequality

Understand the benefits, costs, risks and limits of economic growth

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.

1.7. Interventions

Consider the long-term consequences of interventions that may differ substantially across moral theories

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.

Stakeholder(s):

Longtermists

1.8. Forecasting

Forecast the long-term future

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.

1.8.1. Capability

Consider the ability to forecast the long-term effects of actions

1.8.2. Improvement

Consider how the ability to forecast the long-term effects of actions could be improved

1.8.3. Relevance

Consider which type of long-term forecasts are particularly relevant from a longtermist perspective

1.8.4. Expected Value

Compare the expected value of different strategies for improving the long-term future

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.

1.8.5. Investment

Determine the extent to which longtermists should invest in better forecasts

2. Prioritisation

Consider theoretical and empirical questions for agents engaged in global prioritisation

Stakeholder(s)

Global Prioritisation Agents

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 Making

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.

2.2. Epistemology

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.

2.3. Timing & Discounting

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?

2.4. Diversification & Hedging

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?

2.5. Cost-Effectiveness

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.

2.6. Altruism

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.

2.7. Coordination

Consider how altruistically-motivated actors should interact, particularly when they have different values or beliefs

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.

2.8. Institutions

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.

Stakeholder(s):

Political Institutions **Economic Institutions**

Appendix A. Additional Research

Outline additional areas of possible research that would further GPI's mission

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. Animals

Promote animal welfare

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.

Economic Institutions

A.2. Welfare

Consider welfare maximisation in terms of moral philosophy

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.

Appendix B. Related Research

Highlight closely related areas of existing academic research

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. Costs, Benefits & Effectiveness

Analyze costs, benefits and effectiveness

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).

B.2. Economic Indices

Develop macroeconomic measures that capture more than GDP

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.

B.3. Ethics & Intergenerational Equity

Consider how 'infinitarian paralysis' can be avoided and intergenerational consequences can be evaluated

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.

B.4. Disagreement

Consider how to act in the face of disagreement among those of roughly equal competence

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.

B.5. Demandingness

Consdider the scope of moral obligations toward global welfare maximisation

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.

Stakeholder(s):

Individuals **Institutions**

B.6. Population Ethics

Understand how to compare outcomes across different groups of individuals

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.

B.7. Risk & Ambiguity

Consider the extent to which risk and ambiguity should be avoided

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.

B.8. Moral Uncertainty

Incorporate moral uncertainty into reasoning to prioritise problems

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).

B.9. Information

Consider the value that additional information can provide to increase chances of choosing better actions

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.

B.10. Evidence

Harness and combine evidence

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.

B.11. Altruistic Decision-Making

Understand the psychological mechanisms underlying altruistic behaviour

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.

Administrative Information

Start Date: 2020-10-31
End Date:

Publication Date: 2022-08-27
Source: https://globalprioritiesinstitute.org/wp-content/uploads/GPI-research-agenda-version-2.1.pdf

Submitter:

Given Name: Owen
Surname: Ambur
Email: Owen.Ambur@verizon.net
Phone:

PDF formatted using TopLeaf XML publisher
www.turnkey.com.au