Draft: The Case for Cause Prioritization as the Best Cause

Table of Contents

Introduction

Summary

If modern donors (or governments) were fully informed, they could probably find altruistic causes that have a significantly higher impact than those they currently support. Moreover, surprisingly little research has been directed at the problem of identifying and comparing the most promising causes. These problems may be neglected because they are simply too difficult. However, there are compelling arguments that more research would in fact improve our understanding, and significantly improve the impact of future altruistic activity. In light of these arguments, it seems that research on cause prioritization may be one of the highest-impact activities for altruists to support, especially for those who are particularly concerned with aggregate welfare. The next stage of our investigation is to examine the potential impact of specific projects within this cause.

Overview

Today, philanthropists, governments, and talented individuals all divert some of their energies towards projects aimed at making the world better, from improving global health, to improving education, to mitigating climate change, to promoting social justice, to contributing to technological development, and so on. There are deep disagreements about which of these causes are most worthwhile.

Some of these disagreements are really questions about fundamental values: for example, if we care more or less about the welfare of future generations, we may become more or less concerned with global warming or poverty reduction. But many of these disagreements are the result of fundamental and often implicit disagreements about empirical facts: what is the long-term effect of climate change, or social injustice, or technological development? How efficiently can our current actions influence global health, or education, or global coordination?

Cause prioritization seeks to answer the question: which available causes are most impactful to support? It requires understanding the long-term impacts of the various projects we are considering, as well as understanding the effectiveness with which we can carry out those projects.

The basic argument for prioritization is relatively simple. We are currently uncertain about the best causes to support, and suspect that there may be large predictable differences between the most effective causes and our current best guesses. Therefore *if* we could make progress on prioritization it would have a large positive impact. Moreover, little high quality work has been done on prioritization to date, and there is a good chance that we can make progress on prioritization relatively efficiently. This suggests that prioritization work might have a very high value. Even if it does not, efforts that clarify the difficulties may still have a high value, by allowing future researchers to carry on the project or confidently turn their attention elsewhere.

If we believed that all altruistic projects were being pursued effectively, we might be unconcerned with cause prioritization: it is surely rational for society to spend *some* effort on a very wide range of problems. Unfortunately, this doesn't seem to be the case. Many of the problems we face are sufficiently large and complex that even if we are willing to make a huge investment we can only address a small fraction of all of the smaller problems of which “improving education” or “ending poverty” is composed. And even for those subproblems we choose to tackle we can pursue only a small fraction of all available strategies. Moreover, it seems there are enough plausible altruistic causes that we can pay significant attention to only a small fraction of them. (Note that there is a distinction between plausible causes which may or may not yet have anyone working on them, and plausible projects which are currently underway. The latter is much less numerous, and one could argue that a large fraction of plausible existing projects receive significant attention.)

We might also be unconcerned with prioritization if we thought that our current state of knowledge was relatively sophisticated, and that our remaining uncertainty about cause prioritization reflected our uncertainty about impractically difficult empirical questions. However, at present there seems to be little evidence for this view. Only a small amount of high-quality research has been done on cause prioritization and most of this research has not been carried out in such a way that others could build on it (instead each philanthropist rediscovers the same basic arguments and shares their conclusions but not their reasoning). The work that has been done suggests that progress is difficult but not impossible, and human history reflects positively on our ability to build up a deep collective understanding of a subject and eventually make headway on hard problems. Moreover, even if we think there is a good chance that this project is difficult or impossible, the possible upside is large enough that we should still pursue it---we should simply try to make this work public and transparent, so that our successors can judge its failures and successes, and can be better informed when making their own decisions.

What do we mean by "cause prioritization"?

Values

Ultimately, cause prioritization aims to evaluate the relative merit of different projects according to some particular values. Here we will focus on aggregate welfare, although similar questions could be posed for other values and many of the arguments we make will not depend on this choice. In particular, we will consider very weak time preferences, such that the welfare of future individuals is weighted essentially equally with the welfare of current individuals (although if populations increase significantly the total welfare of future generations might become more important than the welfare of the current generation). To the extent that a system of values is relatively unusual, the tradeoffs involved in cause prioritization change---fewer people have worked on cause prioritization for those values before, but at the same time fewer people care about the results of cause prioritization and successful prioritization will direct future values. Aggregate welfare seems to be in a sweet spot for cause prioritization. There are many existing donors, policy-makers, and intellectuals who are concerned with pushing policies that improve aggregate welfare, but *long-run* aggregate welfare has been the subject of relatively little serious research.

Scope

Conceptually, cause prioritization consists of two related undertakings: understanding the impact of various measurable intermediates, and understanding how well available projects can achieve those intermediates. For example, in addition to understanding *how* to most effectively improve education (and *how much* we can improve it), we would like to ask: “What is the long-run impact of educational improvement? How does it ultimately translate into things we care about for their own sake? How can we reason about the benefits of projects, and compare them against the benefits of very different projects?"

In fact there is a whole range of questions between these two extremes. The most easily measured intermediates are typically very narrow, such as “insecticide-treated mosquito nets delivered.” There is then a whole sequence of more tenuous causal relationships, leading to reduced disease burden, improved health and economic outcomes, very small changes in global dynamics, and ultimately long-lasting changes in human welfare.

Research on different steps of this process are natural complements. Understanding which problems are most important helps us focus our search for effective solutions, and understanding which problems we can effectively solve lets us focus our investigations into their relative importance. Moreover, making a decision between different projects ultimately requires tentative answers to all of these questions.

This document will focus particularly on the distant end of the causal chain, i.e. "how important is this problem in the long run?", for three reasons. First, this area appears to be the most badly neglected, and what work has been done has not been focused on key questions of interest to altruists. Second, the considerations relevant to this question may help us focus our attention in the future even if the best projects are amongst those that haven't yet occurred to us, and may help guide our search for these as-yet-unidentified projects. Third, this area seems most likely to require specific attention, and least likely to be directly addressed by commonsense intuitions or by experience gained in other projects.

Questions and methods

A wide range of questions might fall under this heading: what happens to the world if you increase economic growth, technological progress, wealth inequality, carbon emissions? What are the determinants of war or social instability? What changes today have a lasting effect on the future? How do characteristics of individuals relate to characteristics of society? What are the effects of different disease burdens, what is the value of biodiversity, how bad are the tails on climate change, what are the largest risks to society, when does government investment crowd out private investment? Many of these questions are the subject of established academic fields, and the methods and styles of reasoning involved in cause prioritization need not be novel.

However, there are some differences in perspective between cause prioritization and typical academic inquiry. Cause prioritization is most advanced by finding robust answers to questions, which often means focusing on the *easiest* problems rather than the *hardest* or most interesting (though finding "easy" problems to bite off is quite difficult, and "easy" here might just mean that one can hope to find a satisfactory answer with dozens of person-years of research). Moreover, much of the work in cause prioritization may require coordination between experts in different fields who understand different problems and dynamics. In some cases it is possible to make direct quantitative comparisons between two different causes, but in many cases such quantitative comparisons are error prone and it is more reliable to make a more direct comparison between the causes by comparing the details of each. In general, cause prioritization refers to a small subset of academic work, suggests a different approach to prioritizing that work, and asks slightly different questions than normal academic inquiry; the unifying distinction is the extent to which cause prioritization is guided by the ultimate goal of making comparisons.

One simple approach to cause prioritization would be to construct a single monolithic model that described a society, within which we could examine the effects of policy interventions or charitable projects. Unfortunately, the empirical record seems to be quite bleak for projects with this level of ambition (cf. environmental science or macroeconomic modeling), and it seems unlikely that the attitude exemplified by this approach will be productive in the near term. Instead, successes in this area seem more likely to involve making "rough and ready" comparisons between causes, extracting subjective impressions and improving our methodologies for managing disagreement, searching for weak approximations, managing qualitative arguments, and similar work.

The possible upside

The best causes are probably not well-understood

Society has a relatively rigorous understanding of the direct impacts of certain health interventions, but for most causes our understanding is much weaker. Even for health interventions, the indirect effects of intervention are very poorly understood and have not been the subject of much serious study. For other causes, like social justice, technological progress, or environmental preservation, the long-term effects of contemporary interventions are extremely poorly understood.

If we insist on supporting causes for which the effects are very easily defined and reasoned about, or for which compelling cost-effectiveness analyses are currently available, it seems overwhelmingly likely that we have taken the most important causes off of the table. We'll give three of the strongest reasons for this view: few causes are well-understood, well-understood causes focus on very direct effects, and we can identify particular promising causes that aren't yet understood.

The **first** and most basic reason for this is that we understand (in this strong sense) only a tiny fraction of causes, and so it seems unlikely that the best causes are amongst those that we understand. The only reason that we might already understand the best causes is if our understanding of a cause is an important input into its effectiveness. For example, it might be that the only way to have a large impact is to try many different approaches and focus on those that are empirically the most effective. However, it appears that many projects seek to improve a particular measurable outcome---education levels, CO2 reductions, gross world product, number of casualties of violence, *etc.*---and by focusing on any of those intermediates we could iteratively improve our approach and discover how to have a high impact. To date we have little evidence that we can't understand other causes *well-enough* to pursue those causes effectively, if we knew that the intermediates we were pursuing were valuable.

A **second** reason that well-understood causes are not likely to be optimal is that those causes focus on the near-term effects of our actions, which are probably swamped by the long-term effects. This is because the great majority of people will probably live in the future, which is likely to be very long and very large. So even a modest effect on the welfare of future generations can easily dominate normal effects on existing people, at least according to an aggregative perspective. We can see that some simple interventions predictably have long-range effects on many future generations---for example, further reducing the (already very small) probability of human extinction or social collapse in a massive global conflict would increase the probability that those future generations can exist at all. These easily identified lasting effects are not necessarily the largest ones, but they serve as a "proof of concept" and encourage us to think about other lasting impacts of our actions.

In light of this, the main impacts of well-quantified altruistic interventions are their lasting impacts on the world. But once we acknowledge that these are the main effects even well-quantified interventions become rather speculative. The extent of their impact depends on the long-range effects of contemporary poverty reduction and human empowerment. And unfortunately these impacts are hardly more well-understood than the long-range impacts of quantifiably better education, reduced carbon dioxide emissions, scientific progress, faster economic growth, etc. In each case there are strong intuitive reasons to suspect the long-range impacts to be positive, but determining *how* positive, and making comparisons between them, is currently a domain for intuitions. The fact that health interventions have a significant *immediate* impact does not help us estimate their long-run impact, and indeed because health interventions *focus* on this immediate impact they may be less effective than work that explicitly focuses on important long-term determinants of welfare.

A **third** reason to suspect that we don't yet understand the best causes is that we can identify *particular* altruistic projects which look likely to have a large impact and which we don't yet understand well. In particular, our understanding of the contemporary world suggests that technological development and public policy both play an important role in shaping the welfare of current people and the trajectory of modern society. So projects which accelerate technological development or change the impact of technological development, or projects that help policy-makers identify and implement effective policies, seem likely to have a very large positive impact. Even if we are narrowly focused on reducing poverty in the developing world, it seems plausible that policy or technology could offer cost-effective routes to do so, and we certainly don't yet have evidence that would dissuade someone who favored such an indirect approach.

There are likely to be large differences between the impacts of different causes

A key reason to be optimistic about cause prioritization is the belief that some causes are likely to have a much higher impact than others. If the best causes are even several times more efficient (on the margin) than our current best guesses, it would imply that further prioritization could still increase the value of future altruistic efforts by several fold. If there were larger gaps, the value of prioritization would be even more pronounced.

A number of arguments support the view that there are large differences between different interventions: in areas where quantitative evaluation is possible we have seen large differences, there is a good chance that the "altruistic market" is not sufficiently efficient to eliminate such large differences when they exist, and there are "crucial considerations" that can easily increase or decrease the value of interventions by several times.

The **first** argument is that we have observed large differences in effectiveness between different opportunities, in cases where we have been able to quantitatively evaluate outcomes. For example, although we have much left to understand about health interventions, it is now clear that there are several order-of-magnitude gaps in effectiveness between different interventions that seem equally plausible a priori. For example, the Disease Control Prevention Priorities identified many large gaps; though the data has serious errors, the overall pattern of widely varying effectiveness appears to be robust. When looking at health charities instead of interventions similar gaps are observed, though they are somewhat smaller and aren't perfectly predicted by the intervention cost-effectiveness estimates (they also depend significantly on organizational fundamentals, quality of execution, and context). Outside of charity we observe similar patterns, where e.g. the great majority of the value of startups in aggregate comes from a small fraction with very high impact.

The **second** argument is that the "altruistic market" appears to be significantly less efficient than ordinary markets, and so many of our intuitions about the similarity of different opportunities should not apply. Without some force ensuring that only the best available opportunities remain---such as market forces, natural selection, deliberate optimization, or something similar---it would not be surprising to see that some opportunities are much better than others.

Of course there are degrees of market efficiency, and this is not necessarily to say that we should expect to find proverbial $20 bills on the ground. But it seems clear that less optimization power has been put into searching for the best altruistic opportunities than into searching for the best for-profit opportunities: the charity sector is much smaller, is characterized by a much greater diversity of motives, tends to invest relatively little in analysis and prioritization (at the behest of donors), and so on. There are other players searching for promising altruistic opportunities, such as government funders deciding what to spend on, but they seem to face an even greater variety of motives and an even weaker focus on long-run aggregate welfare. Consequently, we should be more optimistic that a concerted effort to identify the best causes will turn up outliers. Without strong reasons to expect different causes to be particularly close to each other, the expected gap between different causes becomes quite large under realistic assumptions.

The **third** argument is that there are "crucial considerations" which can easily increase or decrease the value of a cause by a large factor, and our opinions on many of these crucial considerations have changed over the last decades---and should be expected to continue to change as we think more carefully. For example, the relative importance of the long-run impact of our activities has only been recently grasped (the first clear presentations of the case seem to have appeared in the early 2000's, see for example Nick Bostrom), and its consequences for altruistic activity are just beginning to be appreciated. More pragmatically, the effect of aid in different countries depends on economic dynamics in the developing world, and these considerations could easily change the total value of any particular project by several times. Other examples abound: the relative importance of animal welfare, the effectiveness of political advocacy, the compounding or diminishing of social changes in the long run, and so on.

Realistic changes to our views on any of these issues could easily increase or decrease the value of an intervention by a factor of 2 or more. In light of this level of uncertainty it is hard to imagine the scenario in which the effectiveness of different interventions ended up clustering in a narrow band.

Without a better understanding of plausible causes, we are unlikely to focus correctly

Even if there is a large gap between the effectiveness of different interventions, it does not follow that prioritization could improve the quality of spending. It could be that we will manage to pursue the highest-impact causes whether or not we make progress on prioritization. There are three basic scenarios in which this might occur: there aren't too many plausible projects and we can (and should) support all of them even though most will prove to be ineffective, we can already identify the best available opportunities and so don't need to make further progress, or pursuing available projects is the best plausible path towards improving our understanding of those interventions.

The **first** scenario supposes that the number of plausible projects is relatively small compared to the available resources, and that these projects either face diminishing marginal returns or that the effort of prioritization is greater than the gains from focusing on the highest-impact projects. This scenario may seem superficially plausible based on a survey of the contemporary charitable landscape: there are a very large number of donors and non-profit employees with many different interests, and only perhaps a few dozen causes that seem to merit serious attention. As long as we pursue the highest impact opportunities in each category, it seems that we will capture *most* of the total positive impact that we could hope for, even if we never learn which causes are most important.

However, there are two significant problems with this simple story: one is that there are a huge number of problems lumped together under a simple heading like "aid," "education," "social justice," or "technology," and the second is that the existence of unfunded projects in an area may understate the opportunities in that area.

Even if we decided that developing world aid was the best cause, we would still face a choice between improving health, education, agriculture, infrastructure, and so on; and even if we decided to focus on global health we would still have a choice between dozens of diseases and broad categories of diseases; and even if we focused on malaria we would have a choice between many approaches; and even if we focused on developing malaria vaccines we would still have a choice between dozens of technical programs, and dozens of approaches to those technical programs, *etc*. At a very coarse level it looks like existing philanthropic funding might be enough to cover all of the plausible altruistic projects, but on closer inspection we are splitting a trillion dollars of philanthropic funding between millions upon millions of possible projects, and most of them simply cannot receive a reasonable amount of funding.

The second challenge is that “room for more funding” in an area, i.e. the existence of promising but unfunded projects, significantly understates the possible opportunities in that area. Investigating and initiating projects in an area requires significant interest and commitment of resources, especially attention and human capital. Identifying an important and neglected area doesn’t just mean funding projects in that area, it also means creating new projects.

The **second** scenario supposes that we can already identify the best available projects, so that further prioritization would have a modest effect on the quality of spending. There are a number of compelling reasons to be skeptical of this possibility.

One is the significant disagreement (both in stated preferences and actions) about what the most effective projects or uses of public funds are. As in the first scenario, it may superficially seem that only a handful of big altruistic projects have any proponents in the public discourse, but when the options being discussed are at the level of generality of “invest in our schools” or “preserve the environment,” you should expect that zooming in would yield many more layers of disagreement. Anyone claiming to have solid answers to these questions must be claiming to be exceptional.

A more direct argument is that many of the most promising approaches to improving the world rely on somewhat speculative assumptions: the effectiveness of certain kinds of advocacy, particular claims about the future, claims about economic systems and economic counterfactuals, and so on. Unless we claim that questions about the long-term effects of increased immigration (for example) are either settled or unapproachable, it is hard to hold that our current understanding is sufficient to identify the best causes.

The **third** argument accepts most of the case for prioritization, but maintains that the most effective way to make headway on prioritization is to work directly on our current best-guess projects. By observing the impact of those projects we can effectively refine our model of the world. This will allow us to better understand those interventions' impacts, and will suggest new interventions to pursue. This argument is compelling and indeed it is hard to deny that "learning by doing" will be an extremely important part of discovering and developing high-impact altruistic projects. But properly understood, this activity seems to be complementary with cause prioritization.

When we work on a typical cause, such as development aid or environmental advocacy, we are at best able to observe and refine our impact on certain measurable intermediates---the health and well-being of the affected poor, carbon emissions, or whatever else. No matter how much aid we supply or environmental advocacy we do, we will not be able to observe the long-run impacts of these activities or understand their ultimate impacts on the world. Instead we will rely on climate models to predict the impact of carbon emissions, or on economic models, case studies, or intuitions to understand the effects of aid. Over the very long term we will get evidence about the ultimate impacts of these activities, but that evidence will be extremely noisy, and moreover the timescales involved are so long (typically many decades) that it is hard to incorporate into pragmatic decisions about what to do next. We will get better and better at achieving the measurable intermediates we work towards, but our understanding of the total value of those intermediates will not be improved.

The most important output of cause prioritization is an improved sense of the value of these intermediates, and in most cases "learning by doing" is not a substitute for this kind of information.

Why we can do it

So far we have argued that *if* we could efficiently make progress on prioritization, *then* it would have a high impact. But it may well be that understanding the relative importance of different intermediates is essentially impossible, and projects that aim to do so will flounder indefinitely while diverting resources from the problems that matter today. On this view, we might be better served by continuing to use our intuitions and simple informal arguments for prioritizing causes.

I think there are a number of compelling reasons to pursue cause prioritization notwithstanding this concern: (1) because so little work has been done so far it seems likely that further research can significantly clarify our understanding, (2) we can see a number of research programs that could directly produce useful information, and (3) even if more targeted approaches to prioritization failed, if they were carried out openly and transparently they would still create significant value merely by clarifying the difficulties.

Little work has been done so far

The strongest reason to be skeptical of cause prioritization is that it is an ambitious project on which society has not yet made significant progress, despite its natural interest. Indeed, the relative priority of different social problems is one of the most common topics of public discussion and debate. However, it appears that in fact relatively little research has actually been directed at cause prioritization. Moreover, the research that has occurred has certainly not come to a standstill, and in fact seems to have relatively efficiently advanced our understanding. The research that *has* occurred so far can be naturally divided into three categories: *implicit cause prioritization* within normal academic research, a small number of *explicit cause prioritization* projects, and cause prioritization that has occurred privately as an input to decisions by philanthropists and individuals.

Cause prioritization within academia

The great majority of public cause prioritization to date has been carried out within academic research, under a number of different headings. Naturally, most fields concern themselves with a small set of related intermediates, and investigate the relationships between those intermediates, the determinants of those intermediates, and the effects of those intermediates.

For example, a great deal of research has been done on various measures of educational attainment at the population level, on the relationship between those measures (and their relationship to important unobserved parameters), on the determinants of educational attainment, and on the effects of changes in education. Growth theorists have investigated the nature, causes, and effects of growth, climate scientists the causes and effects of climate changes, and so on. Econometricians have investigated the relationship between many economic parameters of interest, such as economic output, basic research, technological development, inequality, education, and public finances. Economists and other social scientists have reasoned abstractly about the kinds of mechanisms that we should expect to play a role in social development and have sought empirical confirmation of those effects.

Academics have also tackled some of the conceptual questions involved in cause prioritization. For example, Martin Weitzman has considered the extent to which practical decisions about questions like climate change should be dominated by considerations of extreme, unlikely effects, and has considered the relationship between risk and time-discounting in these situations. Nick Bostrom and others at the Future of Humanity Institute have explored the implications of contemporary actions for the very far future, and the moral claim that such far-future effects dominate the altruistic impacts of our decisions.

But (1) the total investment in these projects altogether appears to be modest, (2) few of these projects have explicitly focused on the questions that are important for evaluating altruistic projects, or sought the kind of approximate understanding that would allow us to decide where to focus further inquiry, (3) few of these projects have actually engaged in quantitative comparisons of different effects, and so for the most part they have not been forced to confront the problems that arise when attempting to make such comparisons. Despite this, it seems that a large number of important considerations and empirical results have been established by this research tradition, which significantly clarify our modern picture of cause prioritization, and those efforts which have focused directly on cause prioritization seem to have contributed especially efficiently to this picture.

Explicit cause prioritization projects

A number of projects have worked to explicit prioritize projects in very different areas, which often requires explicitly evaluating the importance of different outputs.

Perhaps the most prominent is the **Copenhagen Consensus**, for example see the CC 2012 project here. This research primarily identifies and aggregates relevant work already performed in academia. However, the challenge papers also contribute some novel analysis of the economic impacts of various problems and policy proposals, and the panel of economists convened by the consensus also performed some new and valuable work by evaluating the cases presented in those papers. The CC has done an extremely small amount of work to date (they have commissioned and analyzed perhaps a few hundred "challenge papers" and “perspective papers”), although their work nevertheless represents the most ambitious attempt at prioritization that I am aware of.

The CC focuses on economic impacts and contributions to growth; this is a valuable comparison to make, and focusing on economic impact makes the project more manageable (though still extremely ambitious). But by narrowing its scope in this way the CC may fail to address some of the considerations that are most important for evaluating the long-run impact of problems. As a particularly clear example, a 1% chance of destroying all life on Earth may be orders of magnitude worse than a certain 1% reduction in population and capital stock, but the two effects would have a similar cost on this crude economic analysis. Similarly, we probably shouldn't value lost lives by their economic impact in the near term. The reasonableness of these economic approximations depends on complicated arguments about long-run social effects that the CC does not attempt to address.

Relatedly and unfortunately, general perception of the CC has sometimes been dominated by its judgments regarding climate change. The director of the CC is Bjorn Lomborg, who has publicly taken controversial positions on climate change, and the organization of the CC conferences is often seen as biased against climate change interventions. This issue is particularly unfortunate because the kinds of long-run considerations that bear on the prioritization of climate change vs. near-term economic prosperity are extremely important, while the politicization of this issue seriously jeopardizes the prospects for serious discussion. (To whatever extent that Lomborg is in fact motivated by prejudice this is doubly unfortunate, though I really don't want to take a stand on that question given my ignorance.)

A much smaller project is currently being undertaken at **GiveWell** under the name "GiveWell Labs." GiveWell's primary focus historically has been charity evaluation, and their methodology focuses on establishing a very robust case for an intervention by talking at length with experts, taking a skeptical view, being extremely transparent about their work and justification, and focusing on interventions that are amenable to clear analysis. With GiveWell Labs they intend to apply a similar methodology to more "speculative" causes.

GiveWell is relatively skeptical of cost-benefit analysis, and tends to focus on other considerations like track record, other empirical evidence, and expert opinion. It is not yet clear how these methodological choices will play out in the context of cause prioritization. To date they have published a number of shallow overviews, and have looked in more depth at meta-research and development aid, but have not yet really begun the business of prioritization. GiveWell's inclusion in this list is primarily founded on optimism about this research continuing.

The **Disease Control and Prevention Priorities** project performs a very narrow form of cause prioritization within the domain of global health. Health interventions aim to reduce or ameliorate the burden of particular diseases, while the outcomes we ultimately care about are measures like quality of life or economic development. The DCPP evaluates the impact of various interventions on these measures, as an input into prioritizing particular projects implementing those interventions.

A very small amount of cause prioritization is being undertaken at **80,000 Hours** and a handful of like-minded organizations. Like the work at GiveWell this prioritization work is extremely preliminary, and should not be seriously expected to have yet settled the questions or even made large steps.

The number of explicit cause prioritization projects, and the total investment in those projects, is extremely small; for the most part their outputs seem to be promising and tentative in part by virtue of the minuscule investment to date.

Prioritization by philanthropists / individuals

Philanthropists and large foundations do make decisions about which causes to support, and individuals make decisions about which projects to work on. Many of these decisions seem to be determined by historical contingencies or dominated by personal considerations, and even for those guided by altruistic considerations a relatively small fraction are concerned with aggregate welfare. But some of these decisions are based on evaluating the impacts of the various intermediate outcomes those projects could work towards. The work done by philanthropists and individuals may be a large share of all cause prioritization that has been done.

Unfortunately, this work seems to be done primarily in private; to the extent there is high-quality thinking on these questions, it is very rarely shared, and public discussions rarely move forward our overall state of knowledge or reveal important new arguments. As a consequence, even if many people have thought about these issues, most of them have been covering the same ground over and over again, and our collective progress has been consequently slow.

Foreseeable developments

There seem to be foreseeable developments within cause prioritization that would have a material effect on our ability to make accurate judgments about the relative merit of interventions.

For example, recently it has become clear that contemporary actions can have a massive humanitarian effect via their consequences for the far future, and this creates room for new research to elucidate those effects. Some direct influences, for example by the reducing the risk of extinction are well-understood, but for the most part the indirect influences are not understood. Evaluating the plausibility of various long-range effects of our actions is currently underway and seems likely to yield progress (even if the answer is that there are no predictable positive effects). Evaluating the indirect effect of intermediates like education or economic growth in light of our understanding of these long-range effects is also underway, and seems to be both quite new and relatively promising.

Another example is within more traditional research in economics. There is an old and growing literature on the determinants and effects of faster growth, and there is much room to apply this knowledge to those questions which are most relevant to comparisons between plausible altruistic projects. Moreover, engaging in this project would clarify which questions are needed, and would inevitably shift focus to those techniques which can produce decision-relevant conclusions even in light of massive and unresolvable uncertainty.

A final example concerns the methodological issues which seem to play a significant role in cause prioritization. Simply carrying out research on cause prioritization with an eye towards these methodological issues---managing disagreements, producing a useful record which others can build on, limitations of cost-effectiveness analyses---seems likely to build capacity to make progress, even if the object level work proves challenging. Merely experimenting with different approaches seems likely to yield useful information if done publicly and openly, in light of the small amount of work that has been done to date.

Failing transparently would be a significant contribution in itself

One concern with prioritization work may be that it has doubtlessly occurred before, in many contexts and under many names, and that we today run a serious risk of repeating the exercise *ad infinitum* without making meaningful headway. More precisely, the concern is that even if we cannot identify a large body of "failed" prioritization work, the fact that no such projects have achieved a high profile is significant evidence that they have failed to even get off of the ground or make a large enough contribution that we would know about them. Relatedly, we may simply have a peculiar sense of what would qualify as high-quality cause prioritization, such that there *is* a large body of prioritization work that we simply fail to acknowledge as such.

Certainly this is a possibility that deserves attention and probably explicit investigation. But it appears that to date there are no intellectually substantial, transparent, relatively widely known efforts to identify and prioritize high-impact causes. Thus even a failed project could have a large positive impact if it managed to hold itself to high intellectual standards, achieved a modest profile, and was transparent about its reasoning. Such a project would allow successors to understand the difficulties of cause prioritization more clearly, which would facilitate future efforts or provide adequate justification to abandon a failed approach.

We would expect such a project to have a number of distinctive characteristics that we don't observe in historical projects: open and public engagement with dissenting voices, clear exposition of their reasoning process, correct and clear argument, openness to the full class of interventions and effects that informed critics consider plausible, and identification and aggregation of relevant work that has been done historically. It seems that GiveWell and the Copenhagen Consensus have some of these characteristics to an unprecedented extent, though both have a great deal of room for improvement, and

GiveWell in particular appears to have a realistic prospect of making a huge amount of progress in this direction. It’s not clear whether GiveWell has yet reached the level of sophistication or scope at which we can reliably say that it is historically unprecedented in these respects, but it does seem like this is a significant possibility.