

Understanding Effective Altruism and its Challenges¹

Introduction

There are many problems in the world today. Over 750 million people live on less than \$1.90 per day (at purchasing power parity).² Around 5.5 million children die each year of easily-preventable causes such as malaria, diarrhea, or pneumonia.³ Climate change is set to wreak environmental havoc and cost the economy trillions of dollars.⁴ A third of women worldwide have suffered from sexual or other physical violence in their lives.⁵ More than 3,000 nuclear warheads are in high-alert ready-to-launch status around the globe.⁶ Bacteria are becoming antibiotic-resistant.⁷ Partisanship is increasing, and democracy may be in decline.⁸

Given that the world has so many problems, and that these problems are so severe, surely we have a responsibility to do something about it. But what? There are countless problems that we could be addressing, and several different ways of addressing each of those problems. Moreover, our resources are scarce, so as individuals and even as a globe we can't solve all these problems

¹ Parts of this chapter also appear in "The Definition of Effective Altruism", forthcoming in *Philosophical Perspectives on Effective Altruism*, Hilary Greaves and Theron Pummer, eds., forthcoming and Hilary Greaves and William MacAskill, "Global Priorities: A Research Agenda".

² World Bank Group, *Poverty and Shared Prosperity: Taking on Inequality 2016* (Herndon: World Bank Publications, 2016), ch. 2.

³ UNICEF, "Levels & Trends in Child Mortality", https://www.unicef.org/publications/files/Child_Mortality_Report_2017.pdf.

⁴ John Broome, *Climate Matters: Ethics in a Warming World* (New York: W.W. Norton, 2012); William Nordhaus, *The Climate Casino: Risk, Uncertainty, and Economics for a Warming World* (New Haven: Yale University Press, 2015).

⁵ United Nations, *The World's Women 2015: Trends and Statistics* (New York: United Nations, 2015).

⁶ Kelsey Davenport, "Nuclear Weapons: Who Has What at a Glance", Arms Control Association, January 2018, <https://www.armscontrol.org/factsheets/Nuclearweaponswhohaswhat>.

⁷ World Health Organization, "Antibiotic Resistance", October 2016, <http://www.who.int/mediacentre/factsheets/antibiotic-resistance/en/>.

⁸ Pippa Norris and Ronald Inglehart. *Cultural Backlash: The Rise of Populist Authoritarianism* (New York: Cambridge University Press, 2018).

at once. So we must make decisions about how to allocate the resources we have. But on what basis should we make such decisions?

The effective altruism movement has pioneered one approach. Those in this movement try to figure out, of all the possible uses of our resources, which uses will do the most good, impartially considered. This movement is gathering considerable steam, with over 3,300 people taking Giving What We Can's pledge to give at least 10% of their income for the rest of their lives to the organisations they believe to be most cost-effective,⁹ \$90 million per year moved to GiveWell's top recommended charities,¹⁰ a foundation with expected assets of over \$10 billion joining the fold,¹¹ and a community consisting of thousands of people around the world who are trying to use their time on earth to improve the world by as much as they can.

In this chapter, I introduce the moral framework and methodological approach that effective altruism uses to prioritize causes, charities, and careers, and examine some of the world problems that, on this perspective, appear to be most urgent and important: global health and development, non-human animal suffering, and risks to long-term human survival. I then lay out some of the most important unsolved problems facing the effective altruist project.

What is effective altruism?

As defined by the leaders of the movement, effective altruism is the use of evidence and reason to work out how to benefit others as much as possible, and the taking action on that basis.¹² So defined, effective altruism is a *project*, rather than a set of normative commitments. It is both a research project - to figure out how to do the most good - and a practical project, of implementing the best guesses we have about how to do the most good.

There are some defining characteristics of the effective altruist research project. The project is:

- *Maximizing*. The point of the project is to try to do as much good as possible.

⁹ <https://www.givingwhatwecan.org/>.

¹⁰ GiveWell, "Our Impact", <https://www.givewell.org/about/impact>.

¹¹ <https://www.openphilanthropy.org/>.

¹² This is the definition provided by the Centre for Effective Altruism, and has been endorsed by most of the leading organisations or leaders of organisations in the movement. See Center for Effective Altruism, "CEA's Guiding Principles", <https://www.centreforeffectivealtruism.org/ceas-guiding-principles/>. The definition is accompanied by a set of guiding values that help to show how effective altruism should be applied in practice. Those values are: commitment to others, scientific mindset, openness, collaborative spirit, and integrity.

- *Science-aligned.* The best means to figuring out how to do the most good is the scientific method, broadly construed to include reliance on both empirical observation and careful rigorous argument or theoretical models.
- *Tentatively welfarist.* As a tentative hypothesis or a first approximation, goodness is about improving the welfare of individuals.
- *Impartial.* Everyone's welfare is to count equally.

Effective altruism has been endorsed by thousands of individuals and a number of organisations around the world. These organisations include:

- 80,000 Hours, which provides career advice service for talented young people to help them have a large social impact.
- The Centre for Effective Altruism, which runs conferences and local groups to bring together the effective altruism community.
- GiveWell, which tries to identify outstanding giving opportunities within global health and development.
- Giving What We Can, which is a community of people who have pledged to donate 10% of their income to the charities they believe are most effective.
- The Open Philanthropy Project, which advises the foundation Good Ventures on which grants will accomplish as much good as possible.

As we can see from the list of organisations above, a significant focus for the effective altruism community is how *philanthropy* can be used to do as much good as possible. But this is not a defining feature of effective altruism. 80,000 Hours is focused on how individuals can use their *careers* to do as much good as possible. And the question of what *policies* would do the most good is also of increasing interest within effective altruism.

It is important to distinguish effective altruism from utilitarianism.¹³ It is true that effective altruism has some similarities with utilitarianism: it is about maximising the good, and it is primarily focused on improving well-being, with many members of the community making significant personal sacrifices (such as donating 50% of their income) in order to do more good. Unlike utilitarianism, however, effective altruism does not claim that one must always sacrifice one's own interests if one can benefit others to a greater extent. Indeed, on the above definition effective altruism makes *no* claims about what obligations of benevolence one has. Nor does effective altruism claim that all ways of helping others are morally permissible as long as they help others the most; indeed, there is a strong community norm against promoting or engaging in activities that cause harm.¹⁴ Furthermore, all plausible moral views, and not just utilitarianism, recognize that well-being is at least part of the good and that there are *pro tanto* reasons to promote the good.¹⁵ So effective altruism is a project that should be of interest to those from a wide variety of moral perspectives.

How important, then, is the question of how to do the most good with a given unit of resources? The argument that this question is of great importance is based on the fact that different ways of doing good differ *vastly* in the amount of good that they can do. This is counterintuitive: on average, people think that the best charities differ in their effectiveness compared with typical charities only by a factor of 1.5 or so.¹⁶ That is, we naturally think that cost-effectiveness among

¹³ Unfortunately, many critics do not distinguish effective altruism from utilitarianism. John Gray, for example, refers to 'utilitarian effective altruists', and in his critique does not distinguish between effective altruism and utilitarianism. John Gray, "How & How Not to Be Good", *The New York Review of Books*, May 21, 2015, <http://www.nybooks.com/articles/2015/05/21/how-and-how-not-to-be-good/>. Giles Fraser claims that the 'big idea' of effective altruism is 'to encourage a broadly utilitarian/rationalist approach to doing good.' Giles Fraser, "It's Called Effective Altruism—But is it Really the Best Way to Do Good?", *The Guardian*, November 23, 2017, <https://www.theguardian.com/money/belief/2017/nov/23/its-called-effective-altruism-but-is-it-really-the-best-way-to-do-good>; see also Marko Bakić, "How Is Effective Altruism Related to Utilitarianism?", Quora, December 30, 2015 ("EA is a particular flavor of utilitarianism"), <https://www.quora.com/How-is-effective-altruism-related-to-utilitarianism>; Iason Gabriel, "The Logic of Effective Altruism", *Boston Review*, July 6, 2015, <https://bostonreview.net/forum/logic-effective-altruism/iason-gabriel-response-effective-altruism>; and Catherin Tumber, "The Logic of Effective Altruism", *Boston Review*, July 1, 2015, <https://bostonreview.net/forum/logic-effective-altruism/catherine-tumber-response-effective-altruism>.

¹⁴ This is emphasised strongly, for example, in an 80,000 Hours blog post by myself and Ben Todd, "Is it Ever Okay to Take a Harmful Job in Order to Do More Good? An In-depth Analysis", 80,000 Hours blog, <https://80000hours.org/articles/harmful-career/>.

¹⁵ As John Rawls commented: "all ethical doctrines worth our attention must take consequences into account in judging rightness. One which did not would be simply irrational, crazy." John Rawls, *A Theory of Justice* (Cambridge, Mass: Belknap Press, 1971), 30.

¹⁶ Robert Wiblin, "Most People Report Believing it's Incredibly Cheap to Save Lives in the Developing World", 80,000 Hours blog, <https://80000hours.org/2017/05/most-people-report-believing-its-incredibly-cheap-to-save-lives-in-the-developing-world/>

charities is *normally* distributed. But contrary to appearances, the distribution is *heavy-tailed*: the most cost-effective charities are *far* more effective than the typical charities; plausibly a hundred times more effective.

This claim is supported by a variety of lines of empirical evidence. Even in the developing world, different programs — each of which we would consider a good use of money — vary by a factor of ten or a hundred. According to GiveWell, donations to the Against Malaria Foundation will provide a benefit equivalent to a year of healthy life (or QALY) for \$100. In contrast, the estimated cost to provide one year of healthy life by treating Kaposi's Sarcoma is \$50,000, a factor of one hundred more.¹⁷ Once we move to international comparisons, the difference between programs is even more extreme. GiveWell estimates that it costs \$3200 to do an amount of good equivalent to saving a life.¹⁸ By contrast, government departments in the US are willing to pay over \$7 million per life for safety-improving infrastructure.¹⁹ This phenomenon isn't limited to health programs. The same holds true among education programs. Among estimates from the Abdul Latif Jameel Poverty Action Lab, the most effective ways of improving test scores with a given amount of money are hundreds of times more effective than typical programs (which would themselves be regarded as highly cost-effective in a domestic setting). And a similar picture emerges if we just compare the relative value of an additional dollar in rich and poor countries. Survey data and economic evidence suggest that life satisfaction is proportional to the logarithm of income: every doubling of income is associated with a constant increase in life satisfaction. Since typical members of a rich country like the US are one hundred times richer than the poorest 700 million people, additional money going to one of these people will do a hundred times as much to improve their well-being as it will to improve the well-being of member of a rich country.

A second distinctive contribution of effective altruism is the idea of *cause-impartiality*: that the choice of cause (that is, the problem that one is trying to solve) that one focuses on should be determined *only* by the amount of good that one can do by focusing on that cause. In contrast, the prevailing view with respect to doing good is that one's choice of cause is a matter of personal preference. On the prevailing view, it's perhaps true that, once one has chosen a cause, one should try to do as much good as possible with respect to that cause (although even this much more modest requirement is sometimes rejected). But whether one should focus on

¹⁷ Dean Jamison *et al.* (eds.), *Disease Control Priorities in Developing Countries*, 2nd ed. (Oxford: Oxford University, 2006).

¹⁸ GiveWell, "Cost-effectiveness Analysis", 'Results' sheet, https://docs.google.com/spreadsheets/d/1KiWfiAGX_QZhRbC9xkzf3I8IqsXC5kkr-nwY_feVlcM.

¹⁹ Binyamin Appelbaum, "As U.S. Agencies Put More Value on a Life, Businesses Fret", *New York Times*, February 17, 2011, <http://www.nytimes.com/2011/02/17/business/economy/17regulation.html>.

education, or the arts, or climate change — there's no objective answer to that question, and the best approach is to work on whatever cause you're most personally passionate about.

Three arguments support the idea that our reasons of altruism require cause impartiality. The first is simply an appeal to cases. Suppose that, as a volunteer doctor in a resource-starved hospital in a poor country, you can do one of two things with your last day of work before you return home. First, you could perform surgery on an elderly man with prostate cancer, thereby saving his life. Or you could treat two children for malaria, thereby saving both their lives.

If you had a personal attachment to the cause of fighting prostate cancer, would that give you sufficient reason to save the life of the elderly man rather than the two children? Clearly not. The importance of saving two lives rather than one, and of saving people who have much more to gain from their treatment, clearly outweighs whatever reason a personal attachment might bring. Yet this is morally analogous to the decisions that we actually face when we try to use our resources to do good.

The second argument is based on the *arbitrariness* of any way of generalizing from a personal attachment to an individual. Suppose that I had a family member who died of prostate cancer. Does that give me an additional reason to support charities that work on prostate cancer? Why would it give me a reason to support charities that fight prostate cancer, rather than all cancer in general, or all non-transmissible diseases? Or all forms of suffering? And if there was some compelling argument that moved from my personal circumstances to a reason in favor of supporting charities that fight prostate cancer, why should that argument not support a *narrower* problem than that? Why should I not support the cause of British people with prostate cancer? Or people who are over 70 and have prostate cancer? In general, there is no non-arbitrary way of delineating causes. And that means that any argument from personal circumstances to partiality among causes will have to pick from one arbitrary delineation over another.

The third argument I call the *double jeopardy* argument. In general and on average, the causes I am most likely to get to know and therefore form personal attachments to are those which are more salient to me: this includes issues in my home country (such as education in my country), issues that have affected me or those close to me (such as cancer), or issues that have received widespread news coverage (such as natural disasters). But this gives me a reason to think that I will be *less* effective by supporting those causes: the more attention that a cause area gets, and therefore the more resources that are devoted to it, the less good additional resources will do.

The idea of cause-attachments puts those who suffer from less well-known problems into a position of double jeopardy. It is already bad luck that, for example, those in poor countries were born into conditions where it is very difficult to make a minimal income. If we think that it's acceptable to be partial to those causes we have a personal attachment to, then we are condoning the idea that, from the fact that someone was born into a poor country, or are

suffering from a cause that does not get much media attention, not only will they almost certainly live in severe deprivation, but they will also not benefit from our philanthropy, because their problems are not salient to those in rich countries who have the resources to potentially help them.

The individuals that we do not interact with — the global poor, non-human animals, and future generations — are precisely those whom we can help the most. And the reason they are so poorly off is because their plight isn't salient to those who can help the most: well-off inhabitants of rich countries. So it seems doubly unfair if the very reason why these problems do not get the attention they deserve (namely, their lack of salience) gives us a further reason to not to focus on them.

Priority-Setting

Suppose we accept the ideas that we should be trying to do the most good we can with a given amount of resources and that we should be impartial among different causes. A crucial question is: how can we figure out which causes we should focus on? A commonly used heuristic framework in the effective altruism community is a three-factor cause-prioritization framework. On this framework, the overall importance of a cause or problem is regarded as a function of the following three factors:²⁰

- *Scale*: the number of people affected and the degree to which they are affected.
- *Solvability*: the fraction of the problem solved by increasing the resources by a given amount.
- *Neglectedness*: the amount of resources already going towards solving the problem.

The benefits of this framework are that it allows us to at least begin to make comparisons across all sorts of different causes, not merely those where we have existing quantitative cost-effectiveness assessments. However, it's important to bear in mind that the framework is simply a heuristic: there may be outstanding opportunities to do good that are not in causes that would be highly prioritized according to this framework; and there are of course many ways of trying to do good within highly-ranked causes that are not very effective. What's more, the framework also does not take into account the issue of *personal fit*, which is particularly relevant for

²⁰ Formally, we can define these as follows: Scale is good done per by percentage point of the problem solved; solvability is percentage points of a problem solved per percentage point increase in resources devoted to the problem; neglectedness is percentage point increase in resources devoted to the problem per extra hour or dollar invested in addressing the problem. When these three terms are multiplied together, we get the units we care about: good done per extra hour or dollar invested in addressing the problem. See 80,000 Hours, "How to Compare Different Global Problems in Terms of Impact", <https://80000hours.org/articles/problem-framework/>.

someone deciding which problem they should contribute their time or devote their career towards. Since skills and experiences vary across individuals, different people may be ideally suited to working on different problems. Thus, what problem a person should work on is often a function not only of the scale, solvability and neglectedness of that problem, but also of the person's personal fit for working on that problem.

With this framework in our mind, we can see why those in the effective altruism community have prioritized some causes over others, looking at global health, factory farming, and existential risks.

Global health

In terms of the above framework, global health does particularly well on *solvability*. Efforts in global health have an impressive track record. For example, the eradication of smallpox in 1973 saved over 60 million lives — significantly more lives saved than if world peace had been achieved in that same time period — yet it cost just \$1.4 billion, suggesting an average cost per life saved of just \$23.²¹

What's more, there is a large body of high-quality scientific evidence, including randomized controlled trials and meta-analyses, assessing the effectiveness of different programs. And, because individuals' bodies are very similar all round the world, scientific trials generalize more easily than they do in areas such as economic empowerment and education.

GiveWell estimates that through donations to the Against Malaria Foundation, one can reliably do as much good as saving a life for around \$3200.²² Other top-recommended interventions include deworming schoolchildren, seasonal malaria chemoprevention, and vitamin A fortification.

Factory Farming

Every year over 60 billion land animals are killed for food, and the vast majority of these spend their lives in factory farms, experiencing considerable suffering.

Whereas global health is notable for its solvability, factory farming is notable for its neglectedness. Despite the size of the problem, in the US only a few tens of millions of philanthropic dollars are donated every year to organizations that focus on improving the lives of farmed animals — 0.01% of total US philanthropy. The amount is tiny even compared to other

²¹ See William MacAskill, *Doing Good Better: Effective Altruism and a Radical New Way to Make a Difference* (London: Guardian Faber Publishing, 2015), ch. 3 and references therein.

²² GiveWell, "Cost-effectiveness Analysis".

animal causes: there are three thousand times more animals in factory farms than there are stray pets, but factory farming gets one fiftieth of the funding.

A particularly effective set of activities in this area are corporate cage-free campaigns: run by organizations like Mercy for Animals and The Humane League, these campaigns try to convince retailers and restaurants to remove eggs from caged hens from their supply chains. They've been astonishingly successful, securing pledges from all of the top 25 U.S. grocers and 16 of the top 20 U.S. fast food chains.²³ The Open Philanthropy Project, which makes grant recommendations to the foundation Good Ventures, estimates that every \$1 spent on corporate cage-free campaigns has spared over 200 hens from cage confinement.²⁴

Existential risks

An existential risk is a risk of an event that would either annihilate intelligent life on Earth or permanently and drastically curtail its potential. It seems that we currently face at least two existential risks: from global nuclear war, which might result in a nuclear winter, and from climate change, which involves a small probability of extreme warming, such as greater than 10 degree warming. When we look to the next century, we should expect to see new risks emerge as a result of the development of new technology. For example, synthetic biology will probably give us the ability to create a virus with a combination of very high lethality, high contagiousness and a long incubation period. If a virus with such a combination were released, it could infect and then kill everyone in the world before we had the ability to develop countermeasures.²⁵ Another example is the development of artificial general intelligence (AGI). A number of experts are concerned that the task of aligning the goals of an AGI with human goals is much more challenging than it appears, and that if an AGI were created with misaligned goals, it might accidentally destroy civilization.²⁶

There are some compelling arguments that such risks are the most important moral issue facing the world today. Not only would an extinction-level catastrophe kill all seven billion people alive today, it would also result in the curtailment of all of humanity's future potential. Plausibly, this would include the prevention of the existence of an astronomical number of people who would

²³ Lewis Bollard, "Why Are the US Corporate Cage-Free Campaigns Succeeding?", Open Philanthropy Project blog, <https://www.openphilanthropy.org/blog/why-are-us-corporate-cage-free-campaigns-succeeding>.

²⁴ Holden Karnofsky, "Worldview Diversification", Open Philanthropy Project blog, <https://www.openphilanthropy.org/blog/worldview-diversification>.

²⁵ Ali Nouri and Christopher F. Chyba, "Biotechnology and Biosecurity", in *Global Catastrophic Risks*, Nick Bostrom and Milan Ćirković, eds. (Oxford: Oxford University Press, 2008), 450-480.

²⁶ Nick Bostrom, *Superintelligence: Paths, Dangers, Strategies* (Oxford: Oxford University Press, 2014).

have otherwise existed, and the loss of all the artistic and scientific accomplishments associated with the long-term survival of human civilization. If we suppose that human extinction is 100 times worse than the deaths of everyone alive today, then reducing the risk of human extinction by just 0.1 percentage point over this century would be as good as preventing the early deaths of 700 million people.

In the nature of this cause area, it's not possible to quantify the amount of good done per dollar for any given program. Some organizations that work on researching or directly addressing existential risks include the Nuclear Threat Initiative, the Johns Hopkins Center for Health Security, the Center for Human-Compatible AI and the Future of Humanity Institute.

Global health, factory farming and existential risks are three causes that those in the effective altruism community think are particularly high priority, and most individuals in this community focus on one of these three areas. But this shouldn't be taken as a complete list of high-priority cause areas. Other priority causes from an effective altruist perspective include US criminal justice reform, immigration reform, fundamental scientific research and macroeconomic policy. And, indeed, the question of what causes are most promising is itself a very promising cause to work on. I have previously suggested the concept of 'cause X': a cause that is at least as important as our current top causes, but which we have so far either dismissed or not conceptualised. It seems likely that such a cause X exists. The ideas that improving the lives of farmed animals or reducing existential risk might be global priorities are only a few decades old, so it would be very surprising if no new causes were uncovered in the decades to come.²⁷ It is likely that we are currently making some grave errors in our understanding, and that if those errors were corrected we would significantly alter our current understanding of what's most important to work on.

Open Research Questions

In this chapter, I've suggested that effective altruism is best thought of as a research project, rather than as a set of normative claims. That research project is still young, and there are many important unanswered theoretical questions. An incomplete overview of these is as follows.

The value of the future. We may define *long-termism* as the view that the primary determinant of the value of the actions we take today is the effect of those actions on the very long-term future. Many in the effective altruism community believe that long-termism leads to the conclusion that we ought to prioritize existential risk reduction. This inference presupposes that the future will be good. But one can at least imagine some scenarios in which we should expect the future to be negative on balance. How should we assess that possibility? What is the likelihood of good

²⁷ See Derek Parfit, *Reasons and Persons* (Oxford: Oxford University Press, 1984), sect. 154.

outcomes versus bad outcomes? And how should we weight very bad futures compared to very good futures - should we treat the best possible future as equally as good as the worst possible future, or should we give more weight to bad scenarios?²⁸

Indirect effects. Effective altruists generally assume that, in evaluating interventions, we should in principle take into account *all* welfare-relevant effects of those interventions—not just direct effects, like the impact on school attendance from deworming school children, but also indirect effects, like the impacts on population size, economic growth, and government activity. But if we should take indirect effects into account, then, ultimately, we need to assess the impact of our actions from now until the end of time. How should we do this?²⁹

Animal welfare. As noted, one distinctive aspect of the effective altruism community is its focus on improving the welfare of non-human animals. But how should we prioritize between interventions that improve human lives and interventions that improve non-human animal lives? And insofar as we prioritize animal interventions, should we focus exclusively on farm animals, or is there a case for focusing on wild animals as well?³⁰

Cross-cause comparisons. Comparative cost-effectiveness analysis is relatively straightforward when the interventions being compared are sufficiently similar, for example two different ways of relieving blindness, or two different ways of increasing the number of years children spend in school. But this is only a very narrow class of prioritization decisions. We also need to be able to compare very different interventions in terms of ‘amount of good done per dollar spent’, for example programs focused on preventing blindness versus programs increasing child test scores. This requires constructing an adequate common measure of ‘good done’, to create a common numerator, in a principled way.³¹

Diversification. What rationales are there, either for the individual or for the effective altruism community, or world as a whole, to diversify across causes or interventions, rather than simply

²⁸ Nick Bostrom, “Astronomical Waste: The Opportunity Cost of Delayed Technological Development”, *Utilitas* 15 (2003): 308-314; Nick Beckstead, “On the Overwhelming Importance of Shaping the Far Future” (PhD diss., Rutgers University, 2013).

²⁹ Hilary Greaves, “Cluelessness”, *Proceedings of the Aristotelian Society* 116 (2016): 311-339; Brian Tomasik, “Charity Cost-Effectiveness in an Uncertain World”, Foundational Research Institute, <https://foundational-research.org/charity-cost-effectiveness-in-an-uncertain-world/>.

³⁰ Tatjana Višak and Robert Garner (eds.), *The Ethics of Killing Animals* (Oxford: Oxford University Press, 2015); Jeff McMahan, “The Moral Problem of Predation”, in *Philosophy Comes to Dinner: Arguments About the Ethics of Eating*, ed. Andrew Chignell, Terence Cuneo and Matt Halteman (London: Routledge, 2015), 268-294.

³¹ Marc Fleurbaey, “Equivalent Income”, in Matthew D. Adler and Marc Fleurbaey (eds.), *The Oxford Handbook of Well-being and Public Policy* (Oxford: Oxford University Press, 2016), 453-475; Benjamin, Daniel, Kimball, Miles, Heffetz, Ori and Nichole Szembrot, “Beyond Happiness and Satisfaction: Toward Well-Being Indices Based on Stated Preference”, *American Economic Review* 104 (2014): 2698-2735.

identifying the intervention with the highest expected cost-effectiveness and supporting that intervention exclusively?³²

Doing good now vs doing good later. If an altruist wants to do good, she faces the question of *when* to do good. With her money, she could donate right away, she could invest the money at a later date, or she could take out a loan in order to give more now. With her time, she could try to get a high-impact job right away, or she could spend time getting further training or credentials, in order to have a larger impact later on. When should a person spend their altruistic resources now, and when should they instead invest their time or money in order to accomplish more later?³³

Epistemological issues. Figuring out how to do the most good is very difficult, and often it seems that subtle differences in epistemology would lead one to quite different conclusions. These include differences in responses to paucity of hard evidence, in level of trust in abstract arguments leading to counterintuitive conclusions, and in the relative weight placed on different types of evidence. How should we weigh evidence from very different sources, such as evidence from rigorous trials against more speculative considerations?³⁴

Moral uncertainty. We should not be certain in any one particular view of the good. Yet different views of the good will differ significantly in their implications for how we should prioritize among different causes. So, plausibly, we should try to incorporate moral uncertainty into our reasoning when we try to work out how to do the most good. It remains underexplored, however, what implications the fact of moral uncertainty has. How do practical conclusions change when we incorporate reasonable moral uncertainty into our analysis?³⁵

Conclusion

Most members of affluent countries have a tremendous opportunity to make a positive difference to the world: as we saw, just a few thousand dollars can save a child's life, or spare

³² James Snowden, "Does Risk Aversion Give an Agent with Purely Altruistic Preferences a Good Reason to Donate to Multiple Charities?", https://drive.google.com/file/d/0B551Ijx9v_RoMzZrOXZvbG9BMzA/; Holden Karnofsky, "Worldview Diversification", Open Philanthropy Project blog, <https://www.openphilanthropy.org/blog/worldview-diversification>.

³³ William MacAskill, "When Should an Effective Philanthropist Altruist Donate?", forthcoming; Owen Cotton-Barratt, "Should I Help Now or Later?", 80,000 Hours blog, <https://80000hours.org/2015/02/should-i-help-now-or-later/>.

³⁴ Rafaela Hillerbrand, Toby Ord and Anders Sandberg, "Probing the Improbable: Methodological Challenges for Risks with Low Probabilities and High Stakes", *Journal of Risk Research* 13 (2016): 191-205; Holden Karnofsky, "Modeling Extreme Model Uncertainty", <https://www.givewell.org/modeling-extreme-model-uncertainty>.

³⁵ Will MacAskill, Krister Bykvist and Toby Ord, *Moral Uncertainty* (Oxford: Oxford University Press, 2018).

hundreds of thousands of chickens from caged confinement, or increase the chance of preserving the future of civilisation for millions of years to come.

But we still don't know how to do the *most* good. The field of effective altruism is still new, and there is much that remains uncertain and underexplored. What's more, this is a rare area where there is a direct line between philosophical research and influential activists trying to make the world a better place. In this regard, philosophy has an incredible opportunity to contribute not merely to our understanding of the world, but to improving it.