Epistemological, Empirical, and Ethical Critiques of Global Ethics and Effective Altruism

Mark Budolfson & Dean Spears

University of Vermont, Philosophy & University of Texas, Economics

Version 1.3[1]

0. Abstract

Effective altruism and arguments in global ethics tend to assume that there is sufficient evidence on balance for concluding that 'effective altruism' charities help the global poor. We catalog a number of different types of publicly available evidence that defeat these inferences. We focus on two kinds of publicly available evidence that constitute the most pervasive defeaters to arguments in global ethics. First, Nancy Cartwright and Angus Deaton have identified a number of reasons why randomized controlled trials don't support the conclusions about the efficacy of interventions that many applied ethicists assume they do. Furthermore, Deaton provided positive evidence that 'effective altruism' charities should generally be expected to be counterproductive on balance. Second, even if we bracket that kind of objection and assume for the sake of argument that these interventions are just as effective as claimed by their proponents, there are still reasons to believe that attempts by single individuals to contribute to these interventions -- for example, by giving donations to charities -- will generally not do any good. As a novel illustration of these two kinds of defeaters, and as a contribution to explaining the fundamental source of these defeaters and their pervasiveness, we identify a number of recurring principal agent problems and other perverse incentives that predict that these defeaters will very often apply. Finally, with that empirical and epistemological critique of much of global ethics and effective altruism in hand, we turn briefly to the purely ethical issues involved, and attempt to more fully articulate a distinctly ethical objection to Singer's argument from Deaton. We draw attention to the fact that this appears to be another thread of Deaton's criticism of global ethics arguments that promote effective altruism charities, and that Deaton's ethical criticism could benefit from more careful articulation, perhaps from those with expertise on the paternalism literature.

[1] Thanks to Elizabeth Ashford, Luc Bovens, Emily Clough, Jonathan Courtney, Seb Farquhar, Iason Gabriel, Hilary Greaves, Michelle Hutchinson, Will MacAskill, Theron Pummer, Ben Sachs, James Snowden, and participants at the University of St. Andrews conference on the philosophical foundations of effective altruism for very helpful discussions.

1. Introduction: Global Ethics Arguments and the Empirical Objection

Global ethics arguments often have the following structure:[2]

> **Ethical Premise**: There is *no ethically relevant difference* between our everyday situation in which we can easily save members of the global poor by donating several hundred dollars to an effective charity, and a straightforward case where it would be obviously wrong not to provide aid:
>
> > *Singer Variant*: The relevant analogy is to POND: At a financial cost of several hundred dollars to you, you can easily save a child from drowning in a shallow pond.
> >
> > *Pogge Variant*: The relevant analogy is to HARM COMPENSATION: At a financial cost of several hundred dollars to you, you can easily save the lives of people who were unjustly harmed as part of a scheme by your family that has unjustly enriched you to an even larger extent.
> >
> > *Other Variant*: The relevant analogy is to some other case X where it would be obviously wrong not to provide aid.
>
> **Conclusion:** Therefore: It is seriously wrong for each of us not to give large sums charities that provide aid to the global poor, or at least do good for the global poor in some comparably beneficial way.

Philosophers generally take the only important question about this argument to be whether there is any ethically relevant difference between our relation to the global poor and our relation to those in need in the straightforward case where it would be obviously wrong not to provide aid.

As a result, philosophers tend not to question the presupposition of the argument that each of us can in fact help the global poor by giving to charity. However, this is exactly the part of the argument that strikes many non-philosophers as mistaken, because many non-philosophers think it is a mistake to believe that giving to charity or any other form of assistance can be expected to do more good than harm on balance for the global poor. As one leading example, this is one objection offered by Nobel

[2] See Singer 1972, Singer 2009, and Unger 1996, Pogge, 2008.

Laureate Angus Deaton to Singer's arguments and to the effective altruism movement more generally that relies on Singer's variant of the argument above.

In response to Deaton's objections, many philosophers believe that skeptics like Deaton about our ability to help the global poor are making a simple mistake: roughly, the mistake of equating charities that aid the global poor with foreign aid programs that send money to support the *governments* of the global poor.

One of the goals of this paper is to help philosophers see that this is a misunderstanding and an underestimation of Deaton's objections, and that he actually offers (along with others, such as Nancy Cartwright) a deeper epistemological critique of reasoning about the efficacy of aid that philosophers rely upon – ultimately, the critique is that even if the evidence cited by philosophers in support of efficacy was entirely correct, it would not follow that such charities do good on balance.

In addition, Deaton offers an empirical argument that the aid given by 'effective altruism' charities is counterproductive because it has similar dynamics in relevant respects to the dynamics present in cases of counterproductive domestic and foreign aid. The result is an argument not only that it *does not follow* that effective altruism charities are effective, but furthermore that we should actually expect them to be *counterproductive*. This argument does not depend on simply equating effective altruism charities with foreign aid, but rather depends on the claim that we should expect mechanisms to cause foreign aid to do harm to also be present in the case of effective altruism charities, and that when these likely unintended negative side effects are properly taken into account, we should expect effective altruism charities to do more harm than good on balance.

In the next section, we explain these objections, and further unpack the epistemological and empirical assumptions of global ethics arguments. After explaining these objections in detail, in the following section we bracket these kinds of objections, and instead assume for the sake of argument that 'effective altruism' charities really are as effective as their proponents claim. Even so, we argue that there are still other reasons to believe that contributions to those charities by individuals like us will generally make outcomes worse rather than better. Taken together, these sections can be seen as identifying two different types of publicly available evidence that defeat inferences by philosophers for conclusions about the efficacy of donations to 'effective altruism' charities. In the final section, we turn to the purely ethical issues involved, and highlight a fundamental objection to Singer's argument and its ethical premise as endorsed by many effective altruists.

2. Epistemological and Empirical Objections: Deaton's Critique

The discussion above suggests that when representing global ethics arguments, it may be useful to highlight the empirical assumption of such ethics arguments that aid does good when directed through effective altruism charities. The following is one straightforward way of highlighting this empirical assumption, cast here in the context of Singer's argument, which we will use in the remainder of the paper for ease of exposition:[3]

> **Ethical Premise**: There is *no ethically relevant difference* between a case, POND, in which you can easily save a child from drowning in a shallow pond at a small financial cost, and CHARITY, in which you can save a child by giving a financially equivalent donation to an effective charity.
>
> **Empirical Assumption**: CHARITY is a situation that we always find ourselves in, because it is always possible for each of us to reliably identify effective charities that we can give to at little cost and thereby save additional lives that would not otherwise be saved.[4]
>
> **Conclusion:** Therefore: It is seriously wrong for us not to give away large sums to effective charities, or do a comparable amount of good in some other way.

As noted above, critics like Deaton insist that the Empirical Assumption is false. In response, many philosophers are apt to think that Deaton mistakenly equates effective charities with foreign aid programs, and illegitimately ignores the impressive evidence for the good done by effective charities. In what follows, we explain why this is a serious underestimation of Deaton's objections.

To appreciate Deaton's position, it is useful to first unpack the argument that is typically endorsed by philosophers for the Empirical Assumption that some charities are a reliable way to do good. This argument is the basis for belief in the Empirical Assumption by the most sophisticated commentators and leading philosophers, and has essentially the following form:

> **Facts about RCTs**: Randomized controlled trials indicate that intervention i adds x quality adjusted life years in the treatment group per unit of intervention (where x is positive).
>
> **Facts about a Particular Charity**: Charity C has overall budget $D and does w units of intervention i [in a number of locations and contexts, many of which are different from those of the RCTs].
>
> **Inference about Positive Good Done by Charity Per Dollar**: Therefore: we should expect Charity C to add quality adjusted life years to the world at a rate of (x*w) / $D.

[3] This representation of the argument is from Budolfson under review b, which proposes an evaluation of the ethical premise. See Singer and Unger ops. cit. for the argument itself.

[4] For example, here is Peter Singer: "Expert observers and supervisors, sent out by famine relief organizations or permanently stationed in famine-prone areas, can direct our aid to a refugee in Bengal almost as effectively as we could get it to someone in our own block" (Singer 1972, pg. 232; Singer 2016, pg. 8).

> **Conclusion about Net Good Done by Charity Per Dollar**: Therefore: we should expect charity C to add net good to the world at a rate of (x*w) quality adjusted life years / $D.
>
> **Conclusion about Good You Can Do By Giving to Charity**: Therefore: each dollar of your donations to charity C should be expected to do good at the rate of (x*w) quality adjusted life years / $D.
>
> Therefore, the **Empirical Assumption** is true.

To begin to understand Deaton's objections, it is crucial to see that he is engaging with the details of this leading argument for the Empirical Assumption, and his criticism is that even if one grants the Facts about RCTs and the Facts about a Particular Charity, nonetheless the Conclusions that support the Empirical Assumption simply do not follow, and on further empirical grounds should be expected to be false in connection with even 'effective altruism' charities. As such, Deaton's position is very far from ignoring the evidence cited by effective altruists – on the contrary, Deaton's argument is that when we think more seriously about the evidence, it becomes clear that the Empirical Assumption is false.

Deaton identifies a number of independent objections to this argument for the Empirical Assumption. The first objection is to the Inference about Positive Good Done by Charity Per Dollar. One way of putting the basic objection is that the inference assumes that an intervention that has success in particular locations and contexts will have the same success when deployed elsewhere. Deaton would argue that this assumption is always present in RCT-based arguments for the effectiveness of charities – including the top-rated effective altruism charities that distribute bed nets, deworming pills, and the like – because charities engage in interventions in locations and contexts that have not been studied by any RCTs, given that RCTs are extremely expensive, and so only one or a very small handful of RCTs at a small number of locations and contexts are the basis for inferences about what will work in other locations, as per the Inference about Positive Good Done by Charity Per Dollar. (Here we stress that all the points we raise here could be equally applicable to other forms of econometric impact evaluation that don't use RCTs. We speak only of RCTs in what follows simply for ease of exposition.)

With this background in mind, Deaton suggests the following four independent objections to the Inference about Lives Saved from Facts about RCTs:

> **Bias**: RCTs have a very significant bias ignored by the Inference about Lives Saved Per Intervention, because they are almost always conducted at locations and contexts that are antecedently expected to display the maximum and greater than average possible effect – and RCTs that do not demonstrate significant effects are ignored. (One obvious mechanism that explains this: if you are an early career researcher whose career depends almost entirely on having a successful RCT, you have incentive to choose a location and context where the RCT can

be expected to display the maximum effect – and this remains true even if you know that location is not representative.)[5]

Non-Generalizability: Even setting aside the preceding, another problem is that interventions are often not generalizable to other populations / locations / contexts. What works in one village might not work in a neighboring village, and it certainly might not work in another region where people have very different customs and societies, or where different background levels of a disease or other problem are present. Instead, the intervention could do harm elsewhere even if it does good in the particular places selected for the RCT.[6]

Reversal in General Equilibrium: Even setting aside the preceding, we often don't have good reason to think an intervention is going to continue to work when scaled up even within the location of the RCT, partly because the equilibrium that results from a very large society-wide intervention of that sort might have very different properties from the one that emerges from a small sprinkling of interventions within the society.[7]

Short-Term Evaluation: Even setting aside the preceding, we often don't have good reason for to think that the relatively short-term effects measured by an RCT capture the long-term effects within the population. For example, an intervention that dumps food or other goods on an impoverished population might have positive results over a short time horizon yet over the longer term that dumping could destroy the local market for those goods, thereby turning what was a short term undersupply into a permanent one.[8]

For a toy example, imagine an intervention that gives randomly selected farmers a particular kind of non-monetary reward insofar as they produce a particular crop that is judged to be desirable (this is the 'treatment' that those randomly selected farmers are given, in the language of RCTs). We can imagine a randomized controlled trial that shows significant gains for the selected farmers when this treatment is applied within a particular region, based on a trial involving only a tiny sliver of the farmers in that region. But note that (Bias) in the real world this RCT is likely to be conducted in a region where the researchers antecedently have good reason to expect atypically good results, (Non-Generalizability) this treatment might well fail elsewhere especially if people don't value the reward or the crop in the same way elsewhere, and perhaps more importantly such an intervention might fail when scaled up and deployed widely because (Reversal in General Equilibrium) giving this treatment to very many farmers (even in the same region) might destroy markets and perversely incentivize production of a much worse

[5] Compare Deaton 2013, pg. 290.

[6] Cartwright 2012, Cartwright and Hardie 2012, Deaton and Cartwright forthcoming, Deaton 2013, pg. 291.

[7] Deaton 2013, pp. 292-3.

[8] Deaton 2015.

portfolio of crops – and so society might be made worse off by the intervention when it is fully 'rolled out'. (For real-world examples, see Hammer and Spears (2016) on an RCT conducted by the World Bank in 2004 on village sanitation in India, or see the debate about whether effective altruist recommendations of deworming charities are based on flawed inferences from RCTs (Humphreys 2015; see also Berger 2015).)

The problem of Short-Term Evaluation is a useful segue into Deaton's next and independent objection that even if we were to set aside the preceding problems and simply grant that the Inference about Positive Good Done by Charity Per Dollar is sound, nonetheless the Conclusion about Net Good Done by Charity Per Dollar does not follow, and instead should be expected to be false in any particular case. The basic problem is that even if the Inference about Positive Good Done Per Dollar were correct, there might at the same time be **unintended negative side effects** that are not taken into account in the analysis – and for two empirical reasons, we should expect these to be present and outweigh the good that is done in any particular case, even in connection with effective altruism charities. The example of dumping food aid in a region in a way that improves short-term nutritional outcomes but makes the overall outcome worse when longer-term effects are taken into account is an illustration of the very realistic way that an intervention could appear effective according to an RCT that only measured one type of (short term) benefit but do more harm than good on balance if all effects are taken into account.

More importantly, beyond the mere possibility of unintended negative consequences outweighing the genuine good that might be done by an intervention, Deaton has an argument that we should expect more harm than good to be done even by effective altruism charities because of the effect of at least one of the following recurring dynamics:

> **Further Empower Oppressors**: Aid often predictably disproportionately benefits the 'oppressors' in a society, thereby further empowering and entrenching their role as oppressors, and thereby preventing exactly the sort of social change that is necessary for the best feasible long-run outcome for the global poor. (This mechanism is analogous to the way that foreign aid is often counterproductive by amounting to a transfer to a literal dictator, or in a more complex case involving, for example a World Bank grant, is a transfer that disproportionately benefits a small group of people in a developing country, each of whom is an oppressive 'high caste' member, oppressive local boss, or at least in the top 5% of the income distribution in that country, etc.)[9]
>
> **Prevent the Evolution of Good Governance, Institutions, and Capacities**: Even in the absence of the preceding dynamic (which when present is an instance of this dynamic), the most persistent problem with aid is that it blocks the evolution of good governance, local institutions, and local capacities – which is what the global poor most desperately need to achieve the best feasible long-run outcome. For example, even if health aid from the Gates Foundation saves just as many

[9] Compare Deaton 2013, pg. 310, 279, 295.

> lives as claimed by an optimistic RCT-based projection, that aid prevents the evolution of genuine local capacity to administer and provide such goods – and the evolution of state capacity in the health sphere might be exactly a place where good governance and capacity must develop as a precondition for the evolution of other good governance, institutions, and capacities in society. In this way again, aid prevents exactly the sort of change that is necessary for the best feasible long-run outcome for the global poor. (This mechanism is somewhat analogous to the way that a domestic charity might intervenes to provide a public good that is temporarily in short supply that ought to be supplied by the state, thereby removing incentives that would otherwise result in the longer term in the state providing that good.)[10]

The strong claim here is that aid to the global poor should be expected to have the negative effect of preventing state and other capacities from evolving over time in the way it otherwise would have, where this negative side effect is greater than whatever good might be done more directly by the aid.[11] Although it is a strong claim, it is a claim about an empirical issue, and Deaton takes the claim to be well supported by a realistic examination of the evidence from development economics and the history of NGO interventions. Deaton takes this to explain why more optimistic views are continually falsified from economists like Jeffrey Sachs, who claim that small investments targeted in the ways they recommend will have a transformative effect on eradicating poverty. As Deaton puts it, "If it were so simple, the world would already be a much better place." (Deaton 2015).

On the basis of these epistemological and empirical objections (and seemingly other ethical objections to intervening in foreign cultures), Deaton generally opposes the recommendations of global ethics arguments to contribute to effective altruism charities, and believes instead that we should invest in changing the *policies* that prevent the world's poor from doing the best they can, particularly agricultural and trade policy. For those willing to make very large sacrifices for the global poor, Deaton seems to recommend a life devoted to changing these policies, or, more radically, he recommends that one move to an impoverished nation, join the struggle of the poor there, and "cast one's lot with them" (Deaton 2015).

Having explained Deaton's objections (and related criticism from Nancy Cartwright and others of RCT-based inference of effectiveness), for the sake of argument we will set aside these basic criticisms in the remainder of the paper, having noted that it is an empirical question for development economics and the like whether they are in fact sound.

The main point of this section has been to explain the epistemological and empirical nature of these objections – and to show that Deaton's position is very far from merely equating charity with foreign aid, or from ignoring the evidence cited by effective altruists.

[10] Deaton 2013, pg. 292

[11] Deaton 2013, pp. 292-3.

In setting aside Deaton's objections in what follows we do not intend to dismiss them – rather, we set them aside because if his objections are sound, then it is simply game over for global ethics arguments that conclude that we have an obligation to give to effective altruism charities.

3. The Inefficacy Objection to Individual Charitable Giving

In this section we assume for the sake of argument that effective altruists are entirely correct about the good done by its top-rated charities and the virtues of doing that good. Even if we grant all of these things for the sake of argument, our main point in this section is that it is still dubious that *donations* to those charities from individuals like us would do any good. To connect this to the discussion in the previous section, our argument in this section is that even if we assume, contrary to Deaton's arguments, that Conclusion about Net Good Done by Charity Per Dollar is true, it does not follow that we should accept Conclusion about Good You Can Do By Giving to Charity:

> **Conclusion about Net Good Done by Charity Per Dollar**: The amount of good charity C does on balance is ((x/1)*w) / $D.
>
> **Conclusion about Good You Can Do By Giving to Charity**: Therefore: each dollar of your donations to charity C should be expected to do good at the rate of ((x/1)*w) / $D.

To see the worry that arises about this inferences, it is useful to consider GiveDirectly, which we agree with effective altruists is the best candidate for a reliable way of doing good with one's donations. Even though GiveDirectly is highly rated largely based on RCTs, we believe it might be able to escape the critique of RCTs from Deaton and others articulated above. (One of us might be more optimistic than the other about this, but we adopt an optimistic view at least for the sake of argument in what follows.) This is because we suspect that there are special reasons for thinking that auxiliary claims might be true of the sort necessary for conclusions to follow about GiveDirectly from the relevant RCTs in the way that effective altruist evaluations assume.[12] So, we take GiveDirectly to be something like a 'best case scenario' effective altruism charity.

To see first why GiveDirectly – which provides unconditional cash transfers directly to some of the poorest people in a few poor villages in Kenya – might have special promise of avoiding the RCT critique, note that we arguably have special reasons – or we at least have special hope – for thinking that the alleged efficacy of GiveDirectly based on RCTs is indeed generalizable to other locations and contexts, and we have special reasons for thinking that the 'general equilibrium' effect of massively scaling up

[12] For more on the relevant form of the necessary auxiliary assumptions, see Cartwright and Hardie 2012 and Deaton and Cartwright forthcoming.

GiveDirectly would not undermine its success (even if returns per dollar would diminish to some extent). That is because GiveDirectly depends only on a small number of simple mechanisms (giving people money, in the context of a monitored transfer to a cell-phone linked bank account) that are very well understood, and that we can have special confidence are not location or context dependent for their positive effects, and that we have special reason to think would not change the direction of their effect even when scaled up and rolled out across most of that society. Basically, the idea is that we can have special confidence that giving people unconditional cash transfers improves their wellbeing regardless of location if it is done in the sort of minimally conscientious way that GiveDirectly deploys. Largely for these reasons, the effective altruism community has also largely agreed that GiveDirectly is the benchmark against which we should judge other ways of trying to do good.[13] In addition, direct unconditional cash transfers are among the most non-paternalistic aid interventions imaginable, and so avoid a number of ethical worries that might apply to other highly rated charities.

In light of this, we find it particularly useful to assume for the sake of argument that GiveDirectly is known to escape the RCT critique in these ways. To take this assumption a step further, we find it useful to assume for the sake of argument that all of the substantive claims that effective altruists makes about the efficacy of the operations of a charity like GiveDirectly are known to be true with absolute certainty. We can then ask whether there is any further reason to worry that giving to such a charity is not an effective way of doing good.

It turns out that there is. A real-world illustration arises from the fact that several billionaires closely follow the recommendations of effective altruists, and make commitments to 'top up' the revenues of that handful of charities in order to ensure that those charities meet operating budget targets (provided they engage in good faith fundraising and other activities). These are real facts that are readily knowable based on public information described in more detail below. Furthermore, because the amount of plausible shortfall in all of the top-ranked effective altruist charities *combined* is only several tens of millions of dollars per year, and because this group of billionaires have made commitments and have the capacity to top up such charities to erase *much more* than that level of shortfall, in light of this information, we submit that the expectation associated with donations to these charities of, say, a magnitude of $1,000 is simply that slightly less money is transferred from the billionaires to those charities, and that the charities still end up with the same revenues and operating budget.[14]

[13] As leading EA proponent Will MacAskill puts it, "Because cash transfers [of the sort employed by GiveDirectly] is such a simple program, and because the evidence in favor of them is so robust, we could think about them as like the 'index fund' of giving" (MacAskill 2015, pg. 115). This mirrors a similar posture adopted by the leading EA charity evaluator GiveWell, and also adopted by many other EA advocates (Karnofsky 2015b).

[14] In fact, it is not clear that the expectation is that anything happens, as e.g. it might be knowable that donations less than or equal to $1,000 amount to insignificant digits in all of the relevant decision-making by charity organizations, billionaires, and others. We ignore this complication in what follows. For further discussion, see Budolfson forthcoming.

So, contrary to the claims of effective altruists, even assuming the truth of everything that effective altruists assume about the efficacy of the interventions of their top charities, at best the effect of an ordinary individual's donations to their top charities is to transfer money to a billionaire, with no other positive effects – and that is not a wealth transfer that increases the amount of wellbeing in the world.

We emphasize that this is not an argument that charities like GiveDirectly should not exist or should pursue other activities; is not an argument that GiveDirectly's leaders should not spend their time on the organization; and is not an argument that policy-makers at an organization such as the World Bank should not allocate funds to similar cash transfers. Each of these is beyond the scope of this paper, which is focused narrowly on the question of what a normal individual person has reason to do, or may have an obligation to do. Again, we assume for the sake of argument that GiveDirectly and other such charities do exactly as much good through their overall activities as effective altruism (EA) evaluators report. (See Haushofer and Shapiro forthcoming for leading arguments supporting this claim).

What is at issue here is different: the issue here is the expected good done by an additional donation to these charities by an ordinary individual. Our point is that even if these charities do exactly as much good overall as EA evaluators report, the marginal effect of additional donations from individuals is likely zero.

The mere possibility of this dynamic of billionaires standing ready to top up top-rated charities is occasionally acknowledged by effective altruists, but is then quickly dismissed as not relevant to reality, with something like the thought that *of course* this situation does not actually obtain, and merely represents something like a science-fiction possibility.[15] However, on the contrary, we believe that at least something like this situation is often actual, not merely possible. And one of us believes that we can know that this situation actually reliably obtains with respect to all of effective altruism's top-rated charities, and that this makes the expected effect of contributions from normal people to all those charities *negative*, given that a donation then only has the effect of subtracting wealth from normal people like us while increasing the wealth of rich people and their foundations – which, again, is the opposite of a welfare improvement or a better state of affairs in any other respect.

Perhaps the quickest way into the relevant facts is by noting a leading example of the recent 'pivot toward billionaires' that has happened in fundraising for EA charities, together with the uncontroversial fact that there are only a handful of top ranked EA charities, each of which has a surprisingly low limit by its own lights to the resources it can absorb and genuinely turn into welfare gains. The best example of the pivot toward billionaires is provided by Good Ventures, a foundation run by billionaires Cari Tuna and Dustin Moskovitz, which has so much money to invest that it cannot even find nearly enough opportunities to invest its vast resources consistent with the EA criteria it endorses as a constraint on

[15] For example, here is MacAskill: "If DMI will close their funding gap regardless of whether you donate to them [perhaps due to billionaires ready to top them up to budgetary targets], then your specific donation will do very little" (MacAskill 2015, pg. 119). He then simply asserts that this "doesn't pose a problem" given the actual facts, and pays no further attention to the issue (pg. 119). But he does not explain why he is confident that this does not pose an actual problem, and as far as we can tell his confidence is simply misplaced for the reasons we give here.

making donations. In light of this, it reliably fills any genuine need for resources that could be converted into welfare gains by top EA charities.[16]

To more deeply understand the way the pivot toward billionaires has undermined the marginal effect of individual donations, it is important to see that Good Ventures *alone* now represents over two thirds of all money moved to EA charities, as tracked by EA advisor givingwhatwecan.org (MacAskill 2016). Beyond this, the most important facts here are that (a) Good Ventures represents *only one* among a growing number of EA-focused 'billionaires',[17] (b) Good Ventures now bankrolls and colludes with the dominant EA advisor GiveWell in all investment and strategic decision-making, and so Good Ventures makes its decisions in a way that perfectly tracks the dominant EA consensus,[18] (c) Good Ventures *alone* has so much money that, by their own lights they can *by themselves* easily meet all of the funding needs of *all* of the charities that are deemed to be sufficiently effective to be worthy of investment on EA grounds, while still not being able to spend *nearly* as much money as they would like because they judge that after their investments there are no more good EA opportunities for them to invest in,[19] (d) Good Ventures has publicly committed to meeting all the funding needs of the top-ranked EA charities, insofar as those funding needs are connected to those charities actual activities of doing good.[20] Given these

[16] For more detail on all of this – including both the pivot toward billionaires and the surprisingly low limit of extra resources that can effectively be turned into welfare gains by top EA charities – a good place to start is two blog posts from GiveWell, Karnofsky 2015b, and Hassenfeld and Rosenberg, and one blog post from Good Ventures, Karnofsky 2015a. A slightly older but important discussion superseded by the preceding is Karnofsky 2014.

[17] For example, the Effective Altruism Global 2015 conference was advertised as "the largest ever convening of thought leaders, entrepreneurs, billionaires, CEOs, investors, and scientists, and more who are applying reason and data to tackle the world's biggest challenges." In a promotion before the conference, the grand prize was advertised as "1 lucky person will win a ticket to EA Global (Effective Altruism Global) featuring Elon Musk". (Josh Jacobson, "Announcing the Doing Good Better Giveaway", Effective Altruism Forum, online at http://effective-altruism.com/ea/kn/announcing_the_doing_good_better_giveaway, accessed 8 April 2016.)

[18] For example, a post on the Good Ventures website outlining their big picture strategy by the director of both GiveWell and the Open Philanthropy Project begins by stating that "Throughout the post, 'we' refers to GiveWell and Good Ventures, who work as partners on the Open Philanthropy Project", Holden Karnofsky 2015a, accessed 8 April 2016. As a result, we here sometimes use 'GiveWell' to refer to what are, on paper, two organizations, GiveWell and the Open Philanthropy Project, but those two organizations are composed of and managed by the same people, and there is no important wall between their operations relevant to any issue discussed here.

[19] From the Good Ventures blog: "Good Ventures hopes to give away several billion dollars over the coming decades, which – when accounting for likely investment returns – would imply hundreds of millions of dollars per year in grants for an extended period of time at peak giving. In 2014, Good Ventures gave ~$15 million to GiveWell's top charities and an additional ~$8 million based on Open Philanthropy Project recommendations. In other words, their current level of giving is nowhere near where they hope it will eventually be"(Karnofsky 2015a) accessed 8 April 2016.

[20] For both of these aspects of their strategy, see Karnofsky 2015b, accessed 8 April 2016. It is important to note that only a part of what GiveWell calls "room for funding" represents a need for funds that would have an

publicly available facts, we should expect that among charities that are judged by EA to be top charities, any would-be shortfall in donations that in the judgment of EA would have any actual important impact on the operations of that charity will be offset by funding from Good Ventures *alone* – which is, again, only one among a growing number of deep pockets that are closely following EA advice.

In light of these facts, together with other publicly available facts about the decision-making strategy of Good Ventures, it seems that an individual should expect that their donation to these charities will not do any good for the global poor and, at best, will only reduce the amount that billionaires give to these causes, increasing the bank account balances of these billionaires or their foundations.

In response, it could be argued that even granting the points above, nonetheless if individual donations do in fact reduce the amount that billionaires must top up the top-rated EA charities, then that means that those billionaires will then donate the money saved to the next best charities instead – thereby ensuring that an individual's donation does have some significant positive marginal effect.

However, the key claim made by this response is empirically false, as is indicated by further publicly available information that shows that EA-directed billionaires such as Good Ventures are not in fact willing to redirect excess money to charities 'further down the list'. On the contrary, reports commissioned by Good Ventures itself indicate that insofar as individual donations reduce the amount that Good Ventures gives to the top ranked charities, this has zero impact on Good Ventures giving to the next charities down the list, because Good Ventures simply does not then give to the next charities down the list, because it has a policy of only giving to top rated EA opportunities. Good Ventures explicitly endorses the strategy of not redirecting money to charities further down the list, because it believes that the next charities down the list are not worth giving to.[21]

In sum, Good Ventures independently decides what charities deserve what amounts of funding on EA grounds, where this involves only a handful of charities that deserve money in this judgment. It then ensures that their welfare-generating activities are fully funded given the actions of other donors. Thus, it does not have a fixed amount that it independently decides to give with the strategy of giving 'down the list' of top charities until that amount runs out. To think that it does is to endorse a claim about the activities of Good Ventures that is incorrect according to publicly available information released by Good Ventures and GiveWell.

The upshot is that even if we assume that effective altruism charities do exactly as much good as their proponents claim, nonetheless individual donations to those charities still arguably do no good, and in

important impact on those charities actual activities of doing good – for discussion of this, see Hassenfeld and Rosenberg 2015.

[21] See Hassenfeld and Rosenberg 2015, especially the chart summarizing recommended grants by Good Ventures, and the following announcement (Tuna 2015) of Good Ventures actual grants, which tracked these recommendations.

fact do harm insofar as one agrees that a transfer from ordinary people to billionaires is harm – especially when those ordinary people are misled about the nature of the transfer they are making.[22]

4. Correct Consequentialism and Marginal Effect vs. The Average Effect Metrics of Effective Altruism

The preceding section showed that Conclusion about Good You Can Do By Giving to Charity does not follow from Conclusion about Net Good Done by Charity Per Dollar – and provided a strong case for thinking that in fact you cannot do any good by giving to effective altruism charities even if those charities are just as effective as their proponents claim. To see at a deeper level where this reasoning goes wrong, consider the following metric, which is widely used, at least implicitly, by many effective altruism charity evaluators and proponents of the movement:

$$\left(\frac{\Delta Lives\ Saved}{\Delta Donation}\right) = \left(\frac{Total\ Lives\ Saved}{Total\ Budget}\right) \qquad \text{(EA1)}$$

This equation can be read as "the change in lives saved per change in donation to a charity x equals the total lives saved by x divided by the total budget of x in dollars". The idea of this equation is that the marginal effect of, say, a $1,000 contribution to a charity x is equal to: $1,000*(the total lives saved by x / the total budget of x in dollars). For example, here is Will MacAskill:

> The first step in estimating cost-effectiveness is to find out how much the charity spends per person to run their program. For example, it costs about six dollars to deliver one antimalarial bed net, which on average protects two children for two years, so it costs $1.50 to protect one child for one year. It costs GiveDirectly one dollar to give someone in extreme poverty ninety cents; it costs DMI between forty and eighty cents per listener per year to run their education campaigns (MacAskill 2015, pg. 111).

With this sort of evaluation in mind, he writes

> If you donate $1,000 to the Against Malaria Foundation, they will buy and distribute one hundred and sixty bed nets. If you donate $100,000 to the Against Malaria Foundation, they will buy and distribute ten thousand six hundred bed nets (MacAskill 2015, pg. 81).

And here is how effective altruist evaluator GiveWell puts it:

> Historically, GiveWell has sought evidence-backed, thoroughly vetted, underfunded charities for individual donors to support. We've looked for unusually straightforward, evidence-backed value propositions such as "$X delivers Y bednets, which saves Z lives."[23]

[22] For one way of developing a fairness-based objection to effective altruism on this sort of grounds, see Gabriel under review.

[23] From http://www.givewell.org/labs, retrieved 20 March 2016. Note also that GiveWell does not go beyond an average effects approach even in more detailed explanations of its approach, e.g.: "…many commonly cited figures

A more detailed analysis might add an additional term that allows such an analysis to be more readily connected to empirical studies:

$$\left(\frac{\Delta Lives\ Saved}{\Delta Donation}\right) = \left(\frac{Total\ Activity}{Total\ Budget}\right) * \left(\frac{Total\ Lives\ Saved}{Total\ Activity}\right) \qquad \text{(EA2)}$$

Using this metric (EA2), the term Total Lives Saved / Total Activity might be investigated with RCTs and the like, and the term Total Activity / Total Budget can be estimated in a straightforward way.

To see the problem with equations (EA1) and (EA2), which might be called “average effect metrics”, we need only note that marginal effect is not the same thing as average effect – where in connection with EA, we are certainly interested in marginal effect, namely, the actual difference that would be made by additional investment in a charity. For a toy example to illustrate the problem with average effect metrics, imagine a charity that is devoted to saving children from drowning in a particularly dangerous pond, which it accomplishes with sophisticated monitoring devices and child-extraction machinery that it operates at that pond. Imagine that the charity does indeed save numerous children from drowning each year, but it is already much more than ‘fully funded’ because of a particularly ingenious advertising campaign devised by one of its directors, and has thus, so to speak, fully captured all of the opportunities for saving lives in its domain of operation because of the ultrareliable apparatus it has already built – so that now additional revenues are only used to throw increasingly elaborate parties for the staff of this charity. In this case, the charity might perform a worthy function and have a good score via (EA1) or (EA2), but yet it is knowable that the marginal effect of additional donations is zero in terms of additional lives saved.

We believe that at its current best, EA often relies on more sophisticated equations than (EA1) and (EA2), where these more sophisticated equations do not simply equate the marginal effect of additional charity with the average effect in the way that (EA1) and (EA2) do. However, we also think it is fair to note that these leading EA evaluators often slip into reliance on average effect metrics even when more sophisticated marginal effect metrics could be used instead. This is true even in their own description of their methodology, where the barrier is particularly low to at least conceptualizing their evaluation in

are misleading because they (a) account for only a portion of costs (for example, citing the cost of oral rehydration treatment but not the cost of delivering it); and/or (b) cite "cost per item delivered" figures as opposed to "cost per life changed" figures (for example, equating bed nets delivered with deaths averted, even though there are likely many bed nets delivered for each death averted). The cost-effectiveness estimates we use avoid these problems: Our estimates are given in terms of life impact, and specify what sort of life impact can be expected. We generally make at least some attempt to convert impact into units of disability-adjusted life-years (DALYs), a common metric in public health, though we do not always find these units helpful or make them a key input into our recommendations. Estimates include all direct costs associated with interventions, from all involved funders. Thus, planning costs, management costs, distribution costs, etc. are accounted for. Estimates are generally based on actual costs and actual impact from past projects, to the extent this is possible; when we make projections, we attempt to gather all the information we can to inform such projections.” From http://www.givewell.org/international/technical/criteria/cost-effectiveness, accessed 20 March 2016.

terms that correctly invoke marginal effect rather than erroneously equating marginal effect with average effect.

GiveWell is generally taken to have the current best practices for EA evaluation of charities by advocates of EA, to the point that now its evaluations are used by many other leading advocates of EA who previously did their own research. As a result, we believe it is particularly appropriate and fruitful to focus on GiveWell in the remainder of our discussion here, because we also agree that it has best practices among existing EA evaluators. GiveWell can be charitably understood as aiming to use the following more sophisticated metric in its evaluations:

$$\left(\frac{\Delta Lives\ Saved}{\Delta Donation}\right) = \left(\frac{\Delta Activity}{\Delta Budget}\right) * \left(\frac{\Delta Lives\ Saved}{\Delta Activity}\right) \qquad \text{(EA3)}$$

In this equation, the (marginal) effect of a donation is understood as the change in activity (e.g. change in number of bednets distributed) per change in budget (at the margin) multiplied by the change in lives saved per change in activity (at the margin). This equation is on the right track because it invokes actual elasticity terms on the right hand side of the sort relevant to marginal effects, which is an improvement over the explicitly average effect metrics of (EA1) and (EA2).[24]

However, in practice GiveWell's evaluations often invoke estimations and reasoning about the elasticity terms on the right hand side of (EA3) that make their actual method in practice better represented by equation (EA2) above. This is true, for example, as GiveWell often relies only on average effect metrics such as the total activity of an organization divided by its total budget as a proxy for the marginal effect of additional lives saved per additional budget. This reliance on average effect metrics even extends, at times, to the solicitation of subjective judgments from its staff about average effect metrics, which are then used in its evaluations as if they were marginal effect estimates – when solicitation of marginal effect judgments would be more appropriate, and is not made any less feasible by any practical considerations.

In any event, even if one assumes for the sake of argument that EA is entirely correct about the terms on the right hand side of (EA3), and is thus entirely correct about the good done by its top-rated charities, it is often still dubious that *donations* to those charities would do any good for the reasons explained in the previous section, because it seems likely that the donation elasticity of activity may still be zero, because zero might be the correct value of the *Δ budget / Δ donation* term in the correct consequentialist marginal effect equation:

$$\left(\frac{\Delta Lives\ Saved}{\Delta Donation}\right) = \left(\frac{\Delta Budget}{\Delta Donation}\right) * \left(\frac{\Delta Activity}{\Delta Budget}\right) * \left(\frac{\Delta Lives\ Saved}{\Delta Activity}\right) \qquad (\textbf{Correct Consequentialism})$$

[24] In an actual calculation, the activities of a particular charity would be disaggregated based on the portfolio of activities that that charity actually engages in (which, as GiveWell notes, is often surprising). Then, for each activity, one can construct an instance of the right hand side of (EA3) with a subscript i for that activity, and then sum over all instances of the right hand side for all i, yielding a value of $\Delta Lives\ Saved$ / $\Delta Donation$ consistent with the general idea of equation (EA3).

This point remains even when one properly acknowledges the use in EA recommendations of notions such as *crowdedness*, *tractability*, and *impact*.[25] We agree that these concepts are a small step in the right marginalist direction, but they do not make much progress on mitigating the central worry here about the possibility that the donation elasticity of activity might be zero regardless of how much good is done by the charity in total.

To verify that we are not being uncharitable or misunderstanding EA analyses, the reader can examine the actual spreadsheets that GiveWell uses in its charity recommendations, which are available at the following site: http://www.givewell.org/international/technical/criteria/cost-effectiveness/cost-effectiveness-models. Note that despite public discussions of *crowdedness*, *tractability*, and *impact*, those notions do not play much of a role on the spreadsheets – and even if they were incorporated into the spreadsheets fully, they would not resolve the problem that the donation elasticity of activity might be zero. Finally, notions of *crowdedness*, *tractability*, and *impact* are in any event highly imperfect proxies for the marginalist notions they are intended to track, as one of us argues in another paper (Budolfson under review a).

The underlying problem for effective altruism recommendations can be understood as this: existing EA evaluations of charity at best ignore the leftmost term on the right hand side of the equation – Δ Budget / Δ Donation – and at best derive their evaluations entirely based on the rightmost two terms. However, as the phenomenon of billionaires standing ready to top-up charities illustrate, the expected change in budget per change in donation by normal people can be zero, which would make the expected marginal effect of their donation zero even if the rightmost two terms look very good.

In addition to highlighting a flaw in existing effective altruism evaluations of the good individuals can do by giving to charity, the equation Correct Consequentialism above makes more general progress by providing a fundamental consequentialist analysis of the dynamics relevant to the effects of donations to charity.

5. Theoretical Arguments that Δ Budget / Δ Donation is Often Low: Principal-Agent Problems

Moving beyond the argument in the preceding section, the economic literature on charities and non-profits already contains theoretical reasons to suspect that (Δ budget / Δ donation) is often less than 1. Typically, these focus on a cross-donor dimension of crowd-out: if I observe that you donate more, or that the government has allocated more to an organization, then I would rationally conclude that the optimal amount for me to donate is less than it would have been had you not donated (Warr, 1982). This crowd-out of policy and of donation could be perfectly complete, so the marginal effect of a donation is zero (Bernheim, 1986).

[25] See for example https://www.givingwhatwecan.org/research/methodology, http://www.openphilanthropy.org/research/our-process

Although we recommend that the EA movement attend to this existing theoretical literature, in this section we propose a theoretically novel mechanism for donation crowd-out: the principal-agent problem of an organization's fundraising. Principal agent problems arise in microeconomics when principals can only imperfectly monitor agents' efforts – which is almost always the case in actual organizations.

In any sufficiently large development organization to be a candidate for GiveWell's attention, a managerial *principal* who is responsible for the overall direction of the organization is likely to cooperate with *agents* in the organization of multiple types: at least two types are program implementation agents and fundraising agents. It is a special property of international charities, unlike many businesses, that implementation and revenue-collecting agents can be different people, perhaps located on different continents and never encountering one another in person.[26]

In international development, the principal-agent challenges for implementation agents are well-known and well-studied (Chaudhury, et al. 2006; see also World Bank 2004). Indeed, because implementation principal-agent relationships are often a part of the program design being evaluated as a part of a development project, the EA movement explicitly considers these relationships in selecting projects and they are at the heart of the public advocacy by proponents of evidence-based development policy (Banerjee and Duflo 2012).

In contrast, the fundraising principal-agent problem receives little attention in the development economics literature, and almost no attention in the EA literature. However, agency problems may be at least as important in fundraising.[27]

In many charities, fundraising is done by dedicated staff who report to organization principals. Fundraisers are in some way incentivized to successfully raise funds. This incentive could take various forms:

- **Fixed target.** Fundraisers are paid a salary that is independent of the amount of money they raise, except that they are fired if they do not raise enough funds in a specific period.
- **Flexible target.** Fundraisers are paid a fixed salary, and the probability of being fired is decreasing in the amount of funds they raise.
- **Sharecropping.** Fundraisers "sharecrop" with the charity, keeping a fixed percentage of the funds they raise.
- **Billionaire's charade.** A billionaire has promised to ensure the fundraising operation meets the principal's target budget; the fundraising continues merely to save the billionaire some money and to preserve the appearance of a normal charity.

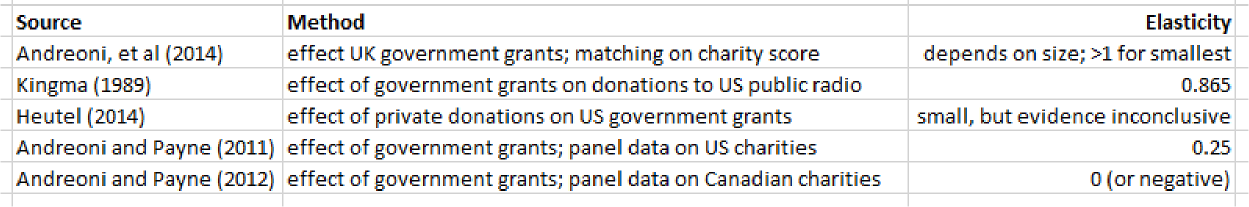
[26] Contrast this with the case of a retail business that is paid precisely when it provides a service to its customer, so fundraising and service provision are necessarily linked.

[27] Perhaps more so, if implementation agency problems are in part resolved through intrinsic pro-social motivation of implementation employees to deliver benefits, which may (or may not) be less likely to motivate fundraising staff.

The consequences of the principal-agent arrangement for effective altruists depend on its details. For example, in the sharecropping case, the elasticity of the organization's budget with respect to a donation is less than one by the amount of the sharecropping. In the fixed target case the elasticity is zero: if effort is costly, then (abstracting away from risk aversion) fundraising agents will always collect precisely their target, and a surprise donation will be entirely captured by the fundraiser in the form of reduced effort, with no extra money passed on to the organization. This implies that in the fixed target case the marginal benefit of a donation in terms of lives saved is precisely zero, no matter how effective the organization's program is at its development goals. This is clearly also true in the billionaire's charade case.

6. Empirical Arguments that Δ Budget / Δ Donation is Often Low

An existing but young empirical literature has estimated the value of Δ Spending / Δ Donation for a number of different kinds of charities. Naturally, like any set of empirical studies, this literature contains research of varying persuasiveness and immediacy of application to the elasticities that EA evaluators need to know. The table below presents a set of estimates from the literature.



Source	Method	Elasticity
Andreoni, et al (2014)	effect UK government grants; matching on charity score	depends on size; >1 for smallest
Kingma (1989)	effect of government grants on donations to US public radio	0.865
Heutel (2014)	effect of private donations on US government grants	small, but evidence inconclusive
Andreoni and Payne (2011)	effect of government grants; panel data on US charities	0.25
Andreoni and Payne (2012)	effect of government grants; panel data on Canadian charities	0 (or negative)

This is not intended as an exhaustive list, nor do we endorse the empirical methods of a paper by including it in this list. In particular, one inapplicability of many of the studies in the table is that they focus on consequences of government grants, rather than small private donations, because large grants are particularly amenable to the techniques of causal identification. These estimates may or may not generalize well to EA evaluation; assessing such generalizability would be an important goal of further investigation.

Despite those limitations, we believe three conclusions are clear from the table:

- Some estimated elasticities are much below 1; these studies therefore give evidence that the problem we highlight exists and could be a large practical concern.
- The estimated elasticities vary radically across studies; these studies do not give us confidence that the elasticity is in fact any particular number.
- Some studies present evidence that the elasticity varies across organizations; this is theoretically expected, and suggests that EA evaluations need organization-specific estimates.

In particular, we note that the empirical literature includes estimates of *zero*. In cases where this is true, the marginal lives saved that would result from a marginal donation would be zero – no matter how effective an organization's programs are and no matter how rigorous and generalizable the evidence of a program's effectiveness is – because the donation would have no effect on the budget or extent of the program implemented. This is not a mere theoretical possibility: it is quantitatively suggested by at least some of the empirical estimates in the literature. If these estimates should be considered wrong or inapplicable, it is important for donors and proponents of EA to know why.

Some studies estimate *interactions*: dependences of funding crowd-out upon other properties of the organization, such as its size. As these interactions become better understood, Give Well could *model* a predicted elasticity for a given organization. We theoretically expect that important dimensions of predicting these interactions could be:

- the nature of the principal-agent relationship between the organization and its fundraisers, that is to say, the type of contract (sharecropping, salary, salary with strict threshold…),
- its size (which influences, for example, its incentive to buy in bulk and make large rather than granular decisions), and
- whether very large donors have guaranteed a minimum budget for the organization.

In sum, EA evaluators could easily incorporate explicit reference to their preferred estimates in these empirical studies or others, just as they currently do for empirical studies about the biological effects of deworming. Recommendations could then be adjusted to appropriately up-weight organizations with large elasticities, such as perhaps small organizations.

Indeed, in the smallest organizations perhaps there is no fundraising principal-agent problem, because the principal is the fundraiser. In such cases, the elasticity could be much greater than one, as an unexpected donation would free up the principal's time to do even more charitable work. These considerations suggest not only that EA's existing metrics *overestimate* the value of donations to large charities, but that EA's existing metrics may also systematically *underestimate* the great value that would be unlocked by donations to many smaller charities.

7: Could field experiments learn the elasticity Δ Budget / Δ Donation?

How could an effective altruist learn the elasticity we need to know, using the principles of evidence-based policy? Consider a field experiment in which a large donor gave $1,200 to each of a set of n randomly selected organizations, out of a larger pool of candidate organizations. $1,200 is a reasonable figure because it is equivalent to $100 a month, a focal donation level discussed on websites such as The Life You Can Save. So, we could imagine 3n candidate organizations organized into n blocks of 3 organizations that are similar on observable characteristics, with one selected for the treatment group and two for the control group out of each group of 3. The experiment would collect organizations' financial disclosure forms for the years before and after the experiment, to see by how much the change over time in the budget of the average organization treated with a $1,200 donation experienced differs from the average change over time in the control organizations. This average change, as a fraction of $1,200, could be taken as an empirical estimate of the elasticity.

Of course, the Deaton critique applies to this field experiment just like any other. This is particularly important because Give Well typically recommends only a few stand-out charities, and this experiment would necessarily average over many, many charities. Additionally, the elasticity from donating $1,200 may be very different from the elasticity of donating much more (or much less); there is no reason to expect it to be linear. Perhaps the blocking could be done in such a way as to allow the experiment to further investigate interactions – such as a larger effect on smaller organizations – but the rules of good experimental practice would require pre-specifying such hypotheses in advance.

The summary statistics from Andreoni and Payne (2011) allow us to consider the sample size (and therefore budget) necessary for such an experiment. In the U.S., 501(c)(3) public charities must disclose their total budget for each year in their form 990. Andreoni and Payne study a set of 8,062 charities which average $797,000 in donations, with a standard deviation of $2,954,000. This large standard deviation indicates the considerable skew in the distribution, with a few large organizations taking in a large total of donations. This large standard deviation is much larger than our potential effect size of $1,200, meaning that it would require an enormous sample (and prohibitively expensive experiment) to be able to at all statistically detect the effect of our donations, in a very simple experiment.

However, the fact that the same organization can be observed over multiple years would considerably reduce the required sample size in the likely event that the variance in organization's year-to-next change in their budget is much smaller than the variance in budget sizes across organizations in a cross section. Additionally, if similar organizations could be successfully blocked into triples before the experiment then our statistical power would be further increased. The precise computations would require statistical 990 data of the form used by Andreoni and Payne.

It may nevertheless be impossible to have a feasibly-sized experiment with sufficient statistical power, if there is simply too much noisy variation in organizations' budgets. One possibility might be to allocate funds differently, making larger (and therefore easier to statistically detect) donations to fewer organizations. This has the potential advantage that the smaller set of organizations could be more similar to the ones Give Well is most interested in, but has the disadvantage that the effect may be different for different sizes of donations, so this would be less representative of the size of donation given by the median Give Well consumer.

Such trade-offs make clear that, just as in the case of program effectiveness, it would be important to think carefully about exactly what average effect it is important to estimate, and why it might generalize. It may be impossible to generate evidence about the elasticity of budgets with respect to marginal donations of the rigor and persuasiveness that EA advocates have come to demand of the other terms in the full EA equation. Effective Altruists may unfortunately be unable ever to have confidence about some of the weakest links in their chain.

8. What is Deaton's Purely Ethical Objection to Singer and Effective Altruism?

In addition to the epistemological and empirical objections described above, Deaton at times seems to have a purely ethical objection to aid, even conditional on aid doing good on balance. For example, Deaton says in critiquing effective altruism:

> …why do the world's poor have such a passive role in all this happiness creation? Why are they not asked if they wish to participate…? Singer does nothing to persuade us that they have volunteered to be the objects of the 'effective' altruism he endorses; indeed, Gallup and Afrobarometer polls show that Africans' own priorities lie elsewhere.[28]

But it is difficult to articulate this objection exactly. One initially tempting way of understanding Deaton's purely ethical objection here might be as analogous populist objections to the power structure involved in the global neoliberal world order: the objection is that it is wrong for a system to be set up so that people in developed nations largely determine outcomes and the rules of the game for people in developing nations contrary to the wishes of the latter. These questions dovetail with more general questions about social policy, and the legitimacy and desirability of international institutions, and the 'democratic deficit' that arguably exists whereby global elites or at least rich nations control most of these institutions – which, consistent with that criticism, might also arguably be the best feasible institutions for promoting geopolitical stability and global wellbeing.

However, this populist view prioritizes local ownership and control over wellbeing in a way that Deaton would presumably find implausible: should people remain starving and in poverty rather than cede some control to outsiders and the global marketplace?

The evidence from Deaton's writings suggests that does not give priority to these other factors over wellbeing, and in fact he favors global trade together with improvements in the international institutions that currently make global trade unfair – but he does not suggest that the status quo is worse than a protectionist alternative.

So, Deaton does not endorse a familiar anti-neoliberalism. At the same time, he does seem to think that there is something deeply offensive about giving people aid without their consent, perhaps especially when they have deeply help values and a culture different from our own. This is, it seems, the central kernel of his distinctively ethical objection, distinct from his earlier arguments that combine ethical and empirical points. But how does this avoid overgeneralizing to imply not only that we should get rid of noxious intervention, but implausibly that we should get rid of most free trade and globalization even if it is highly welfare improving simply because it undermines local decision making power in a way that local populist movements increasingly resent?

We are not sure how to further articulate Deaton's objection here in detail. Our goal is merely to draw attention to the fact that this appears to be another thread of his criticism of global ethics arguments that promote effective altruism charities, and to make clear that his ethical criticism could benefit from more careful articulation. We suspect that others with more expertise in the paternalism literature can

[28] Deaton 2015.

identify a number of nuanced positions that are good candidates for filling in the details of Deaton's ethical objection here – as well as the best objections and replies to those positions.

9. Conclusion

We began by explaining the leading epistemological and empirical critique of philosophers' reasoning about the efficacy of aid. The critique from Nancy Cartwright, Angus Deaton, and others is that even if the evidence cited by philosophers in support of efficacy was entirely correct, it would not follow that the relevant charities do good on balance. In addition, Deaton offers an empirical argument that the aid given by 'effective altruism' charities is counterproductive because it has similar dynamics in relevant respects to the dynamics present in cases of counterproductive domestic and foreign aid. The result is an argument not only that it does not follow that effective altruism charities are effective, but furthermore that we should actually expect them to do more harm than good on balance.

Then, we showed that even if one assumes that all of those objections are mistaken, and that effective altruism charities do exactly as much good as their proponents claim, nonetheless a normal individual should expect his or her contributions to those charities to do no good, based on the fact that billionaires stand ready and able to meet any funding needs of effective altruism charities. This led to our consequentialist analysis of the marginal effect of donations (and, more generally, the marginal effect of any kind of investment in altruism), from which it is clear that the existing principles and metrics endorsed by effective altruists focus on only one part of this equation. When other parts of the equation are considered, it follows that donations that score very well on the existing metrics endorsed by effective altruism often appear to have zero marginal effect (or even negative effects).

So even if one assumes for the sake of argument that Deaton's critique is wrong and effective altruists are entirely correct about the good done by their top-rated charities and the virtues of doing that good, it is still dubious that *donations* to those charities would do any good. By clearly distinguishing the factors relevant to impact, the marginal effect equation we articulate also helps to clarify the logical space of factors relevant to the evaluation of charitable investments, as well as the logical space of objections to the effectiveness of specific charities. With this analysis in hand, effective altruists can improve their principles and metrics in light of all this, including improving investments in research intended to yield insight into how to do the most good.[29]

Finally, we turned briefly to Deaton's distinctively ethical worries. However, apart from identifying a basic kernel of anti-cross-cultural-paternalism, we are not sure how to correctly articulate Deaton's distinctively ethical objection.

[29] We discuss more concrete ideas for how effective altruists might improve their evaluations in Budolfson and Spears unpublished and Budolfson under review a.

[END]

Bibliography

Andreoni, J. and A. A. Payne. 2003. Do government grants to private charities crowd out giving or fund-raising?. *American Economic Review*: 792-812.

Andreoni, J. and A. A. Payne. 2011. Is crowding out due entirely to fundraising? Evidence from a panel of charities. *Journal of Public Economics* 95.5: 334-343.

Andreoni, J. and A. A. Payne. 2012. *Crowding-out charitable contributions in Canada: New knowledge from the north*. Working Paper No. w17635. National Bureau of Economic Research.

Andreoni, J., A. A. Payne., and S. Smith. 2014. Do grants to charities crowd out other income? Evidence from the UK. *Journal of Public Economics* 114: 75-86.

Banerjee, A. and E. Duflo. 2012. *Poor economics: A radical rethinking of the way to fight global poverty*. PublicAffairs.

Berger, A. 2015. New Deworming reanalyses and the Cochrane Review. Online at: http://blog.givewell.org/2015/07/24/new-deworming-reanalyses-and-cochrane-review.

Bernheim, B. D. 1986. On the voluntary and involuntary provision of public goods. *American Economic Review*: 789-793.

Budolfson, M. under review a. Utilitarian Virtues of Boring Low-Hanging Fruit, Even When Investing Many Millions: Suggestions for Good Ventures, GiveWell, and Other Effective Altruism Philanthropy Advisors.

Budolfson, M. under review b. Global Ethics and the Problem with Singer and Unger's Argument for an Extreme Duty to Provide Aid.

Budolfson, M. forthcoming. "The Inefficacy Objection to Consequentialism, and the Problem with the Expected Consequences Response. *Philosophical Studies.*

Budolfson, M. and D. Spears, "Effective Altruism, Marginal Impact, and Fundraising: Weak Links in Effective Altruism's Efficacy Chain", unpublished.

Carey, R. ed. 2015. *The Effective Altruism Handbook*. Centre for Effective Altruism.

Cartwright, N. 2012. Will this Policy work for You? Predicting Effectiveness Better: How Philosophy Helps. *Philosophy of Science,* 79 (5) pp. 973-989.

Cartwright, N. and A. Deaton. unpublished, paper on RCTs.

Chaudhury, N., et al. 2006. Missing in action: teacher and health worker absence in developing countries. *The Journal of Economic Perspectives* 20.1: 91-116.

Deaton, A. 2015, Response to Effective Altruism. *Boston Review*, online athttp://bostonreview.net/forum/peter-singer-logic-effective-altruism

Deaton, A. 2013. *The Great Escape*. Princeton UP.

Duncan, B. 2004 A theory of impact philanthropy. *Journal of Public Economics*, 88, 9–10: 2159–2180.

Gabriel, I. under review. "Is Effective Altruism Fair to Small Donors?"

Gabriel, I. 2016. Effective Altruism and its Critics. *Journal of Applied Philosophy.* Page references are to the 'online first' edition.

Gates, B. 2014. *The Great Escape* is an Excellent Book With One Big Flaw. Online at https://www.gatesnotes.com/Books/Great-Escape-An-Excellent-Book-With-One-Big-Flaw

GiveWell.org. 2015. Spreadsheet Methodology. Online at http://www.givewell.org/files/DWDA%202009/Interventions/GiveWell_cost-effectiveness_analysis_2015.xlsx

Hammer, J. and D. Spears. 2016. Village sanitation and child health: Effects and external validity in a randomized field experiment in rural India. *Journal of Health Economics*.

Hassenfeld, E. and J. Rosenberg. 2015. Our updated top charities for giving season 2015. Online at http://blog.givewell.org/2015/11/18/our-updated-top-charities-for-giving-season-2015

Haushofer, J. and J. Shapiro. forthcoming. The Short-term impact of unconditional cash transfers to the poor: experimental evidence from Kenya. *Quarterly Journal of Economics*.

Heutel, G. 2014. Crowding out and crowding in of private donations and government grants. *Public Finance Review* 42.2: 143-175.

Karnofsky, H. 2015a. Should the Open Philanthropy Project be Recommending More/Larger Grants?. Online at http://www.goodventures.org/research-and-ideas/blog/should-the-open-philanthropy-project-be-recommending-more-larger-grants

Karnofsky, H. 2015b. Good Ventures and Giving Now vs. Later. Online at http://blog.givewell.org/2015/11/25/good-ventures-and-giving-now-vs-later

Karnofsky, H. 2014. Donor coordination and the 'giver's dilemma'. Online at http://blog.givewell.org/2014/12/02/donor-coordination-and-the-givers-dilemma

Humphreys, M. 2015. What Has Been Learned from the Deworming Replications: A Nonpartisan View. Online at http://www.columbia.edu/~mh2245/w/worms.html

Kingma, B. 1989 An Accurate Measurement of the Crowd-Out Effect, Income Effect, and Price Effect for Charitable Contributions. *Journal of Political Economy* 97.5: 1197-1207.

Lichtenberg, J. 2013. *Distant Strangers*. Cambridge UP.

MacAskill, W. 2015. *Doing Good Better*, Guardian Faber.

MacAskill, W. 2016. Presentation at Yale University, 6 May 2016.

MacFarquhar, L. 2015. *Strangers Drowning*. Penguin.

McGoey, L. 2015. *No Such thing as a Free Gift*, Verso.

Pogge, T. 2008. *World Poverty and Human Rights*, second edition. Polity.

Singer, P. 1972. Famine, Affluence, and Morality. *Philosophy & Public Affairs.*

Singer, P. 2009. *The Life You Can Save*. Random House.

Singer, P. 2015. *The Most Good You Can Do*. Yale UP.

Singer, P. 2016, *Famine, Affluence, and Morality*. Oxford UP

Tuna, C. 2015. Our Grants to GiveWell's 2015 Recommended Charities. Online at http://www.goodventures.org/research-and-ideas/blog/our-grants-to-givewells-2015-recommended-charities.

Unger, P. 1996. *Living High and Letting Die: Our Illusion of Innocence*. Oxford UP.

Warr, P. 1982. Pareto optimal redistribution and private charity. *Journal of Public Economics* 19.1: 131-138.

World Bank. 2004. *World Development Report: Making Services Work for Poor People*. World Bank.