William MacAskill misrepresents much of the evidence underlying his key arguments in "Doing Good Better"

−1

guzey

Nov 17, 2018

7 comments

Conflicts of interest: in 2018 I unsuccessfully applied for Grants Evaluator position at the Centre for Effective Altruism and for Research Analyst position at Open Philanthropy.

Whenever possible I use Internet Archive to link to sources, dated at MacAskill's citation date of them.

See /r/slatestarcodex discussion here and /r/effectivealtruism discussion here.

Crossposted from my site.

Background

William MacAskill is an Associate Professor of Philosophy at Oxford University and a Senior Research Fellow at Global Priorities Institute. He co-founded Giving What We Can, Centre for Effective Altruism, and 80,000 Hours. He is one of the founders of the Effective Altruism movement.

His book, *Doing Good Better: How Effective Altruism Can Help You Make a Difference*, published in 2015, was endorsed by Steven Levitt, Steven Pinker, Peter Singer, Julia Galef, and Tyler Cowen (IA, archive.fo). It appears to be endorsed by the Centre for Effective Altruism (IA, archive.fo).

I read this book on the basis of a recommendation by my best friend, who is an Effective Altruist. I have a habit of checking the sources of the books that I read. With *Doing Good Better*, this habit very quickly started to bring fruit. I found that MacAskill regularly misquotes and misrepresents his sources. Eventually, the

volume and egregiousness of misrepresentations reached a point where I decided it wasn't just carelessness or a result of the book being adapted to a broader audience.

Summary

In this post, I show that in *Doing Good Better*, William MacAskill repeatedly misrepresents his sources and uses these misrepresentations to advance the book's key conclusions.

I violate the EA norm of arguing against arguments and not people. I argue that MacAskill is not just being careless with facts and interpretations, but that he is acting in bad faith, as evidenced by

- the volume of serious misrepresentations
- the fact that MacAskill knew for many years that some of the arguments he makes are wrong an/or dishonest

The Centre for Effective Altruism has emphasized (archive.fo) the importance of honesty for the community. I conclude that the Effective Altruism community and the Centre for Effective Altruism must seriously consider whether MacAskill adheres to the standards they set for their members.

In the process of my research for this post, my confidence in truthfulness of GiveWell has increased; of GiveDirectly — has decreased; of MacAskill — has decreased significantly.

My confidence in deworming benefits has decreased.

I show that:

1. MacAskill applies different criteria to different programs while directly comparing them and arguing that one is better than the other. He argues against distributing textbooks on the basis of absence of effect on test scores, but he ignores the absence of effect on test scores for deworming, which he compares to distributing textbooks.
2. MacAskill selectively reports the findings that support his argument from a deworming paper he cites. He writes about increased income, increased working hours, increased tax revenue. He doesn't mention absence of impact of deworming on hemoglobin level, grades of schooling attained, test scores, etc.
3. MacAskill interprets GiveWell's cost-effectiveness estimates literally. For example, he writes that GiveWell's Against Malaria Foundation (AMF) estimate is "most rigorous" and that £20,000 is enough to save ten lives on the basis of it.
    - GiveWell wrote about this estimate specifically that interpreting it literally is wrong and that they provide it for comparative purposes.
    - In 2011, GiveWell wrote a blog post in which they argued that cost-effectiveness shouldn't be interpreted literally, using a quote by MacAskill as an example of that, and in the comments section of the post wrote to Giving What We Can's founder Toby Ord (MacAskill is a

co-founder) that it seems that MacAskill interprets cost-effectiveness estimates literally.

4. **MacAskill introduces the "100x Multiplier", despite knowing that in 2011 GiveWell found one estimate (DCP2 deworming estimate), on the basis of which he previously made a "100 times more effective" argument, to be optimistic by a factor of about 100.**
    - As far as in 2014, Giving What We Can continued to rely on these DCP2 estimates.
5. MacAskill rhetorically asks whether not taking any salary as a charity's CEO means the charity is amazing, implying that this is what Charity Navigator's views lead to. MacAskill provides a quote from Charity Navigator's site as a demonstration of its position. The very first sentence that follows MacAskill's quote from Charity Navigator's site demonstrates that he misquotes and misrepresents Charity Navigator's position.
6. MacAskill writes that income increases happiness solely based on a correlation between income and happiness, despite later in the book writing "Of course, correlation is not causation".

1, 2 are important, because deworming is one of the causes MacAskill recommends the most and it's a central case study that he uses to advance the EA style of program evaluation.

3, 4 are important, because the literal interpretation of these estimates underlies MacAskill's key arguments about which charities to donate your money to and which career to choose to spend your life on. Furthermore, it shows that MacAskill has continued to misrepresent GiveWell for many years, knowing that he misrepresents them.

5 is important, because it demonstrates that MacAskill misquotes his sources even in cases when it's impossible to do so by accident and it is most curious, because the misrepresentation doesn't even help any of MacAskill's key arguments.

I discuss 6 and several other, harder to explain or less egregious, misquotes and misrepresentations in Addendum.

Educational benefits of distributing textbooks and deworming

Summary of this section

MacAskill contrasts the effects of deworming to other programs, while applying different criteria to them and then proceeds to cherry pick the most favorable results from the paper he cites.

MacAskill (please see appendix for full context):

> ICS had been trying to improve school attendance and test scores. ...

> With the help of collaborators, Kremer tested the different ICS programs one by one. First, he looked at the efficacy of providing schools with additional textbooks. Classrooms would often have only one textbook for a class of thirty, so it seemed obvious that providing more textbooks would help students learn. However, **when Kremer tested this theory by comparing test scores between schools that received books and those that didn't, he found no effect for all but the most high-achieving of students. [emphasis mine]**

Let's pause here for a moment. We see that distributing textbooks (we'll call it *Program A*) is not very useful — it has no impact on learning, instrumented for by test scores, for most students. But suppose, there is some program B that was also evaluated and for which we have the data about its impact on test scores and its impact on school attendance. If we wanted to compare Program A to Program B, how would we do it?

It seems clear that if we have the data about the impact each program has on test scores, we should compare them. And, in addition to that, we also may look at attendance. Let's return to MacAskill:

> **Absenteeism is a chronic problem in schools in Kenya, and deworming reduced it by 25 percent.** In fact, every child treated spent an extra two weeks in school, and every one hundred dollars spent on the program provided a total of ten years of additional school attendance among all students. Enabling a child to spend an extra day in school therefore cost just five cents. **It wasn't merely that deworming children "worked" at getting children into school. It worked incredibly well.**
>
> What's more, deworming didn't merely have educational benefits. It had health and economic benefits, too. Intestinal worms can cause a variety of maladies, including anemia, intestinal obstruction, and a suppressed immune system that can increase the risk of other diseases like malaria. Deworming decreases all these risks. **[emphasis mine]**

Notice the substitution: when MacAskill wants to show no effect, he uses test scores; when he wants to show effect, he uses school attendance. If test scores do matter, it's dishonest to not mention them for deworming; if they don't matter, it's dishonest to use them against distributing textbooks. You could suspect that we simply don't know the impact of deworming on test scores, but this is not the case. The deworming paper MacAskill cites (Miguel and Kremer, 2004) here *did* evaluate the impact of deworming on test scores — MacAskill just chooses not to tell us about it, selectively picking metrics that support his conclusion:

> ... The **program reduced school absenteeism** in treatment schools by one-quarter, and was far cheaper than alternative ways of boosting school participation. ... Yet **we do not find evidence that deworming improved academic test scores. [emphasis mine]**

MacAskill uses absence of effect on test scores as evidence against distributing textbooks, but does not use absence of effect on test scores as evidence against deworming.

Later in the book (in chapter three), MacAskill again uses school attendance to show how effective deworming is, while not mentioning the absence of effect on test scores:

> deworming schoolchildren does fifteen times better than that [providing free school uniforms], with 139 total years of school per $1,000.

Cost-effectiveness estimates

Summary of this section

1. In 2010, MacAskill interpreted cost-effectiveness estimates literally.
2. In 2011, GiveWell wrote a blog post in which they argued against using cost-effectiveness estimates literally and used MacAskill's use of cost-effectiveness estimates as one of the examples of that.
3. In 2011, GiveWell found DCP2 deworming estimates to be optimistic by a factor of about 100.
4. In 2014, GiveWell published their cost-effectiveness estimate of *Against Malaria Foundation* and wrote "we believe that cost-effectiveness estimates such as these should not be taken literally" one paragraph after the estimate.
5. **In 2014, Giving What We Can flagship page (first in the Key Pages menu) said "even restricted to the field of health programs in developing countries, research shows that some are up to 1,000 times as effective as others", with DCP2 estimates being the reference** (IA, my screenshot of IA, archive.fo).
6. In 2015, MacAskill published a book in which he
    1. Interprets GiveWell's AMF cost-effectiveness estimate literally
    2. Makes "100 times more effective" arguments

MacAskill interprets GiveWell's AMF estimate in a way they specifically wrote not to interpret it

MacAskill writes that $3400 is the cost to save a life in the developing world, based on "the most rigorous estimates":

> What we've seen is that thinking carefully about how you can do the most to benefit others doesn't just allow you to do a bit more good—it enables you to do vastly more than you might have done otherwise.

> Imagine saving a single person's life: you pass a burning building, kick the door down, rush through the smoke and flames, and drag a young child to safety. If you did that, it would stay with you for the rest of your life. If you saved several people's lives—running into a burning building one week, rescuing someone from drowning the next week, and diving in front of a bullet the week after—you'd think your life was really special. You'd be in the news. You'd be a hero.
>
> But we can do far more than that.
>
> **According to the most rigorous estimates, the cost to save a life in the developing world is about $3,400 (or $100 for one QALY [Quality-adjusted life year]).** This is a small enough amount that most of us in affluent countries could donate that amount every year while maintaining about the same quality of life. Rather than just saving one life, we could save a life every working year of our lives. Donating to charity is not nearly as glamorous as kicking down the door of a burning building, but the benefits are just as great. Through the simple act of donating to the most effective charities, we have the power to save dozens of lives. That's pretty amazing. **[emphasis mine]**

The GiveWell page (archive.fo) MacAskill cites for this estimate says:

> **Using $5.30 as the total cost per net in Malawi and $7.50 for DRC, we estimate the cost per child life saved through an AMF-funded LLIN distribution at about $3,340** [81]
>
> This does not include other potential benefits of LLINs (non-fatal cases of malaria prevented, prevention of deaths in age groups other than under-5 year olds, prevention of other mosquito-borne diseases, etc.). Full details at our report on mass distribution of LLINs.
>
> **As a general note on the limitations to this kind of cost-effectiveness analysis, we believe that cost-effectiveness estimates such as these should not be taken literally, due to the significant uncertainty around them.** We provide these estimates (a) for comparative purposes and (b) because working on them helps us ensure that we are thinking through as many of the relevant issues as possible. **[emphasis mine]**

Here are more examples of MacAskill interpreting that estimate literally. Writing about "five hundred times the benefit":

> the QALY allows us to make comparisons across very different programs that combat very different illnesses. By donating to the Against Malaria Foundation, which buys and distributes long-lasting insecticide-treated bed nets, you would, by this estimate, provide **five hundred times the benefit** as

> you would by spending the same amount of money treating Kaposi's sarcoma. **[emphasis mine]**

Writing about $100 per QALY and "five hundred times" more benefit:

> In the United States, public health experts regard any program that provides one QALY for less than $50,000 as a good value, and health programs will often be funded even if the cost per QALY is much higher than $50,000. In contrast, providing the same benefit in poor countries (such as by distributing insecticide-treated bed nets to prevent the spread of malaria) can cost **as little as one hundred dollars.** That means that, with a given amount of money, you can benefit people in poor countries **five hundred times** more than people in rich countries.
>
> Again, we see the 100x Multiplier at work. We're about one hundred times richer than the poorest billion people in the world, and we can do several hundred times more to help them than we can to help others in the rich countries we live in. **[emphasis mine]**

It's clear that MacAskill misquotes GiveWell; and while he is not technically lying — GiveWell did indeed try to produce the most rigorous estimate — he used this estimate in precisely the way GiveWell warned against, by interpreting it literally.

Why am I so sure that MacAskill interprets GiveWell's estimate in a way they don't want it to be interpreted? Maybe they mean something else by "interpreting it literally", than what MacAskill does?

Here's the reason: because in 2011, Holden Karnofsky (a co-founder of GiveWell and the Executive Director of Open Philanthropy) expressed precisely the same concern, although wording it very carefully, about MacAskill's (and GWWC's) use of cost-effectiveness estimates. In the comments section of his *Why We Can't Take Expected Value Estimates Literally (Even When They're Unbiased)* (archive.fo) post, Karnofsky has a lengthy exchange with Toby Ord (a co-founder of the Effective Altruism movement (archive.fo) and the founder of Giving What We Can (archive.fo)) in which he writes:

> **There are several cases in which Giving What We Can appears to take cost-effectiveness estimates literally even though they do not include adjustments of the sort described here. The quote from Will in the above post is one example; [i.e. Karnofsky literally used a quote from MacAskill as an example of a literal interpretation of a cost-effectiveness estimate in the post]** another is the statement that "Charities which focus on STH, schistosomiasis, and LF are at the very top end of cost-effectiveness — **about 100 times more effective** than typical developing-world health interventions." **[emphasis mine]**

And in another comment Karnofsky writes:

> I think it is worth noting that **Will's 2010 comments appear inconsistent with the reasoning of this post (not just neglecting to mention it).** He argues that even a deworming charity known to be wasting 26% of its money "still would do very well (taking DCP2: $3.4/DALY * (1/0.74) = $4.6/DALY – slightly better than their most optimistic estimate for DOTS [$5/DALY]" and concludes that "the advocacy questions [relevant to whether this 26% is spent effectively] don't need to be answered in order to make a recommendation." **This is a much stronger statement than the one you attribute to him and seems to require taking the estimates literally. {emphasis mine}**

So, as early as in 2011 Holden Karnofsky was concerned about MacAskill's literal interpretation of cost-effectiveness estimates and about his use of "100 times more effective" argument.

Here's an entire subsection devoted to that,

MacAskill's 100x Multiplier

During his discussion of life satisfaction and income, MacAskill introduces the "100x Multiplier":

> This idea is important enough that I've given it a name. I call it the 100x Multiplier. **For those of us living in rich countries, you should expect to be able to do at least one hundred times as much to benefit other people as you can to benefit yourself. [emphasis mine]**

And lest you think that he is being metaphorical, he clarified what he thinks about the Multiplier in Notes:

> **Note that the figure of one hundred is a baseline. I believe that if we try hard, we should be able to do even more good for even less personal cost.** This is for two reasons. First, we've only looked at one problem: global poverty. As discussed in chapter ten, there may be even better opportunities for helping others, in which case the 100x Multiplier is an underestimate. **[emphasis mine]**

Later, MacAskill combines his literal interpretation of GiveWell's $3,400 estimate with earning-to-give:

> **Earlier I said that one of the most cost-effective ways to save lives is by distributing antimalarial bed nets: $3,400 pays for 560 nets, which on average will prevent one death due to malaria.** By pursuing medical oncology, Greg could therefore donate 50 percent of his $200,000 per year earnings while still having a very comfortable $100,000 per year pretax salary (donations are tax-deductible). His donations would save dozens of lives a

> year, considerably more than he could have done if he'd worked directly in a poor country. ...
>
> **In 2014, Greg donated £20,000, enough to save ten lives. [emphasis mine]**

Wait. But what if I am wrong, after all, and MacAskill does realize that these estimates are not precise and they are just estimates? Here's MacAskill:

> **Importantly, the cost-effectiveness estimates given are just that: estimates.** The figures for Kaposi's sarcoma, condom distribution, and antiretroviral therapy are individual estimates based on specific contexts and may therefore be optimistic. The figure for bed-net distribution is more robust—the calculation behind it tries to correct for biases in favor of optimism, and takes into account the specific context in which the charities work—but even this estimate should not be taken as gospel. However, in the context of fat-tailed distributions, even rough estimates are vitally important for decision making. **In the health-care graph, the best program is estimated to be five hundred times more effective than the worst program (which, remember, is still a good program). Even if the highest estimates were too optimistic by a factor of fifty, it would still be vitally important to focus on the best programs rather than merely good ones. [emphasis mine]**

He doesn't: he mouths the words but he doesn't believe what he writes. He says that even if the highest estimates are optimistic by a factor of fifty — that's ok, because the best program is five hundred times more effective.

Except that in 2011, GiveWell discovered (archive.fo) that the DCP2 cost-effectiveness estimate for deworming that GiveWell used and on the basis of which Giving What We Can (where MacAskill is a co-founder) recommended deworming (archive.fo) was off **by a factor of about 100.** MacAskill knew about this and even commented on the GiveWell post that announced DCP2 errors.

And in 2014, GiveWell's optimistic estimate for Deworm the World Initiative's cost per equivalent life saved differed from pessimistic **by a factor of 2118** (GiveWell's spreadsheet, local copy). MacAskill references the GiveWell page (archive.fo) that extensively discusses deworming and links to this spreadsheet when discussing impact of deworming on health in the book's introduction.

By the way, in the book, MacAskill writes that the evidence behind *Deworm the World Initiative* is "fairly robust" and that the evidence behind *Schistosomiasis Control Initiative* is "very robust".

Charity Navigator and the final evidence of MacAskill's bad intent

If you are trained in charitable interpretations and steelmanning or have heard of MacAskill previously and know of his status and reputation in the EA community,

you might still be thinking,

> But how can you make claims about MacAskill's intentions? How can you be sure he and his editors are not just dumbing things down for a broader audience and being *really* careless?

This *is* a difficult question. Believing that the other person is dishonest seems like a particularly easy way to fail at a conversation; and attacking MacAskill's motivations may seem like an extremely inconsiderate thing to do. Maybe he's just sloppy. Maybe his editors are incompetent. Why am I so sure he is acting in bad faith?

I believe that I provided enough evidence above. But, if you're still not convinced, I will present another quote from the book, which misrepresents its source so violently, and where the bad intent is so clear, I don't think it's possible to defend MacAskill as simply being careless (even though this quote does not advance any of the book's key arguments).

MacAskill:

> One popular way of evaluating a charity is to look at financial information regarding how the charity spends its money. How much does the charity spend on administration? How much is its CEO paid? What percentage of donations are put directly to the charity's main programs? **This is the approach that Charity Navigator, the oldest and most popular charity evaluator, has taken for the last fifteen years. According to Charity Navigator, "Savvy donors know that the financial health of a charity is a strong indicator of the charity's programmatic performance. They know that in most cause areas, the most efficient charities spend 75 percent or more of their budget on their programs and services and less than 25 percent on fund-raising and administrative fees."**
>
> **Using these metrics, let's see how the three charities compare.**
>
> Books For Africa's overhead costs are a tiny 0.8 percent of their total expenditure (which was $24 million in 2013), and their CEO is paid $116,204, which is only 0.47 percent of that total expenditure. For these reasons, and for their general financial transparency, Charity Navigator has given BFA its highest four-star rating for seven years running. ...
>
> You certainly wouldn't think about how much Apple and Microsoft each spend on administration, and you wouldn't think about how much their respective CEOs are paid. Why would you? As a consumer you only care about the product you get with the money you spend; details about the financials of the companies who make the products are almost always irrelevant. If Apple spent a lot of money to attract a more talented management team, you might even consider that a good sign that their products were the best on the market!

> If we don't care about financial information when we buy products for ourselves, why should we care about financial information when we buy products for other people? **Take a silly example: imagine I set up a charity that distributes doughnuts to hungry police officers and I am so enthusiastic about the mission that I manage to spend only 0.1 percent of the charity's money on overhead, with the rest spent on doughnuts and distribution. Suppose, moreover, that I, as the CEO of this charity, don't take a salary at all. Would I really have created an amazing charity? [emphasis mine]**

According to MacAskill, the metrics Charity Navigator uses lead to an absurd conclusion that 0.1 percent overhead with zero CEO salary lead to a charity being considered amazing.

Here's the full quote from the Charity Navigator's page (archive.fo) that MacAskill cites:

> Savvy donors know that the financial health of a charity is a strong indicator of the charity's programmatic performance. They know that in most cause areas, the most efficient charities spend 75% or more of their budget on their programs and services and less than 25% on fundraising and administrative fees. **However, they also understand that mid-to-large sized charities do require a strong infrastructure therefore a claim of zero fundraising and/or administrative fees is unlikely at best. They understand that a charity's ability to sustain its programs over time is just as important as its short-term day-to-day spending practices. Therefore, savvy donors also seek out charities that are able to grow their revenue at least at the rate of inflation, that continue to invest in their programs and that have some money saved for a rainy day.** All of this analysis is provided on Charity Navigator's website for free, but when considering groups not found here, savvy donors ask the charity for copies of its three most recent Forms 990. Not only can the donor examine the charity's finances, but the charity's willingness to send the documents is a good way to assess its commitment to transparency. ...
>
> **Sophisticated donors realize that charities need to pay their top leaders a competitive salary in order to attract and retain the kind of talent needed to run a multi-million dollar organization and produce results.** But they also don't just take the CEO's compensation at face value; they benchmark it against similar-sized organizations engaged in similar work and located in the same region of the country. To help you make your own decision, Charity Navigator's analysis reveals that the average CEO's compensation of the charities we evaluate is almost $150,000. In general, salaries tend to be higher in the northeast and at arts and education charities. **Sophisticated donors also put the CEO's salary into context by examining the overall performance of the organization. They know it is better to contribute to a**

> **charity with a well-paid CEO that is meeting its goals than to support a charity with an underpaid CEO that fails to deliver on its promises.** ...
>
> **Although it takes some effort on their part to assess a charity's programmatic impact, donors who are committed to advancing real change believe that it is worth their time.** Before they make a contribution, they talk with the charity to learn about its accomplishments, goals and challenges. These donors are prepared to walk away from any charity that is unable or unwilling to participate in this type of conversation. **[emphasis mine]**

MacAskill completely misrepresents Charity Navigator's views on overhead and on CEO pay. The only thing that we need to do to learn that he misrepresents Charity Navigator: read the very first sentence that MacAskill doesn't cite.

William MacAskill on honesty

On 12 January 2017, William MacAskill wrote in a comment to EA Has A Lying Problem (archive.fo) post by Sarah Constantin:

> **I think that the issue of honesty for people who are consequentialist-sympathetic is very important.** Insofar as pure consequentialists don't place any intrinsic disvalue on promise-keeping or honesty, they are likely to be trusted less as a result – which is a very bad thing if you want to do good in the world! **This makes it *extra* important for consequentialist-sympathetic groups to place great emphasis on honesty and promise-keeping,** and try to cultivate personalities where not being honest is very difficult psychologically for them. **[emphasis mine]**

Centre for Effective Altruism on honesty

The page devoted to Centre for Effective Altruism' guiding principles (archive.fo) says:

> **Because we believe that trust, cooperation, and accurate information are essential to doing good, we strive to be honest and trustworthy.** More broadly, we strive to follow those rules of good conduct that allow communities (and the people within them) to thrive. We also value the reputation of effective altruism, and recognize that our actions reflect on it. **[emphasis mine]**

effectivealtruism.org (archive.fo) (run by the Centre for Effective Altruism) lists *Doing Good Better* on their Resources (archive.fo) page and has a page (archive.fo) devoted to the book.

Conclusion

As I wrote in the beginning of this post, I show the following:

1. MacAskill applies different criteria to different programs while directly comparing them and arguing that one is better than the other. He argues against distributing textbooks on the basis of absence of effect on test scores, but he ignores the absence of effect on test scores for deworming, which he compares to distributing textbooks.
2. MacAskill selectively reports the findings that support his argument from a deworming paper he cites. He writes about increased income, increased working hours, increased tax revenue. He doesn't mention absence of impact of deworming on hemoglobin level, grades of schooling attained, test scores, etc.
3. MacAskill interprets GiveWell's cost-effectiveness estimates literally. For example, he writes that GiveWell's Against Malaria Foundation (AMF) estimate is "most rigorous" and that £20,000 is enough to save ten lives on the basis of it.
    - GiveWell wrote about this estimate specifically that interpreting it literally is wrong and that they provide it for comparative purposes.
    - In 2011, GiveWell wrote a blog post in which they argued that cost-effectiveness shouldn't be interpreted literally, using a quote by MacAskill as an example of that, and in the comments section of the post wrote to Giving What We Can's founder Toby Ord (MacAskill is a co-founder) that it seems that MacAskill interprets cost-effectiveness estimates literally.
4. **MacAskill introduces the "100x Multiplier", despite knowing that in 2011 GiveWell found one estimate (DCP2 deworming estimate), on the basis of which he previously made a "100 times more effective" argument, to be optimistic by a factor of about 100.**
    - As far as in 2014, Giving What We Can continued to rely on these DCP2 estimates.
5. MacAskill rhetorically asks whether not taking any salary as a charity's CEO means the charity is amazing, implying that this is what Charity Navigator's views lead to. MacAskill provides a quote from Charity Navigator's site as a demonstration of its position. The very first sentence that follows MacAskill's quote from Charity Navigator's site demonstrates that he misquotes and misrepresents Charity Navigator's position.
6. MacAskill writes that income increases happiness solely based on a correlation between income and happiness, despite later in the book writing "Of course, correlation is not causation".

How can we interpret these charitably?

1. MacAskill (or his research assistant) forgot that the abstract of the deworming study he cites mentioned absence of effect on test scores or didn't think that he should mention them.

2. MacAskill (or his research assistant) didn't have time to look at other outcome variables of the study.
3. MacAskill forgot that GiveWell believes that using cost-effectiveness estimates literally is wrong and that they specifically wrote not to interpret their AMF estimate literally or just thought it wasn't a big deal.
4. MacAskill thinks that today's estimates are more accurate and unlikely to be wrong by a factor of 100.
5. MacAskill (or his research assistant) accidentally stopped reading Charity Navigator's website at the exact sentence that would make his misinterpretation of Charity Navigator impossible.
6. MacAskill believes that when talking about income and happiness correlation implies causation.

Some of these charitable interpretations seem at least somewhat plausible individually. They absolutely don't look plausible together.

The pattern they reveal is of ignoring the criticism. Of cherry picking and misrepresenting the evidence. Of misquoting other people. And of arguing in bad faith.

One thing I'm particularly stupefied by is the wide praise for the book. It came out in 2015. It was endorsed by a range of people I deeply respect. It was positively reviewed by Marginal Revolution (archive.fo), The Guardian (archive.fo), and Quillette (archive.fo).

80,000 Hours, which MacAskill co-founded, was funded by Y Combinator (archive.fo) and received $1,635,000 over 2017 (archive.fo) and 2018 (archive.fo) from Open Philanthropy. (although, Open Philanthropy did have some reservations regarding 80,000 Hours)

I will conclude this post by quoting MacAskill's *Setting Community Norms and Values: A response to the InIn Open Letter* (archive.fo), published on the Effective Altruism Forum in 2016:

> As documented in the Open Letter, Intentional Insights have been systematically misleading in their public communications on many occasions, have astroturfed, and have engaged in morally dubious hiring practices. But what's been most remarkable about this affair is how little Gleb has been willing to change his actions in light of this documentation. If I had been in his position, I'd have radically revised my activities, or quit my position long ago. Making mistakes is something we all do. But ploughing ahead with your plans despite extensive, deep and well-substantiated criticism of them by many thoughtful members of the EA community — who are telling you not just that your plans are misguided but that they are actively harmful — is not ok. It's the opposite of what effective altruism stands for.

Addendum: MacAskill's Errata page

Here's MacAskill's Errata page for the book. I encourage you to check it out for yourself, especially the comments on Deworming (contrast his deworming cost thought experiment with Karnofsky's 2010 blog post) and PlayPump (contrast his comment with Colin Morris' response letter at the end of the page).

Note that that page doesn't address any of my concerns with the book.

Addendum: GiveDirectly

GiveDirectly's "independent evaluation"

MacAskill quotes GiveDirectly as saying (source of the quote):

> An independent evaluation of our work in Kenya by Innovations for Poverty Action found that recipients use transfers for a wide variety of purposes that on average generate large income gains. Common uses range from buying food to investing in tangible assets such as housing and livestock to investing in children's education.

Note that in the book MacAskill links to a 2013 draft of the GiveDirectly's RCT paper (the same one GiveDirectly links in the quote above), which mentions Shapiro's previous affiliation with GiveDirectly on pages 1 and 7. Thus, he either (1) didn't ever open that paper or somehow missed that, and didn't know that Shapiro co-founded GiveDirectly or (2) knew that the evaluation was not done by the "independent development think tank Innovations for Poverty Action" as GiveDirectly page stated.

GiveDirectly's reporting of the results

GiveDirectly (none of the archiving sites save the page correctly for some reason):

> These conversations have demonstrated that we did a bad job in our initial note in describing the data and the range of interpretations one might take from them, including more negative interpretations. We take responsibility for that. The post was criticized for lacking "nuance and detail," and we agree with that assessment.

The post they were responding to:

> "So, what's the problem," you might ask. "You told us all of this in short form in your last post. Why run through it in more detail?" Well, I, along with others, did get some push back on my interpretation that the nine-month impacts are no longer there. In particular, **GiveDirectly got in touch to inform me that I had missed their blog post on HS (18), published on February 14, 2018.** This is true: I had missed it and immediately updated my

> post to set the record straight. Then, **I read their post. I could barely believe what I was reading.** I am pasting a paragraph from it here, but the whole post is short: please read it in its entirety, so you don't have to take my word. …
>
> Contrast this with the abstract of HS (18) above. **Note, in particular, the lack of detail or nuance in the blog post. Whereas the HS (18) abstract mentions every time which estimate refers to what type of comparison, the above paragraph only gives us great news: all effects are sustained; new positive effects appeared; some effects are even larger now!** Sigh… **[emphasis mine]**

Addendum: Animal Charity Evaluators (ACE)

MacAskill:

> According to Animal Charity Evaluators (a research charity I helped to set up), by donating to charities like Mercy For Animals or the Humane League, which distribute leaflets on vegetarianism, it costs about one hundred dollars to convince one person to stop eating meat for one year.

ACE is a profoundly dishonest organization. See: The Actual Number is Almost Surely Higher (Internet Archive).

Also see Concerns with ACE research. Although, note that despite listing concerns, this article is very sympathetic to ACE, which results in rather amusing results. For example, *The Actual Number is Almost Surely Higher* says:

> The research conducted by these organizations [ACE and The Humane League] is not merely unreliable, but systematically deceptive.

While *Concerns with ACE research* says:

> ACE's research has been criticised in the past, most notably in a December 2016 blogpost by Harrison Nathan. ACE's research has improved since then with some of the most serious problems being resolved.

With research being systematically deceptive, apparently, simply being one of the problems.

Addendum: the deworming study of Sarah Baird, Joan Hamory Hicks, Michael Kremer, and Edward Miguel

Suppose, you deworm a bunch of children, later collect the data on them, e.g.

- did their hemoglobin levels increase?
- did they become happier?
- did their wellbeing improve?
- did their health expenditures fall?

But then you find that these variables didn't move in the right direction. What do you do? Do you have to show these variables? Or can you drop them?

The four variables above were reported in May 2011 draft of *Worms at Work: Long-run Impacts of Child Health Gains* and returned the following results (Table 3):

- Hemoglobin (Hb) level (1999, 2001 parasitological survey samples): **no effect**
- Self-reported currently "very happy": **no effect**
- Index of wellbeing (0 to 1): **no effect**
- Respondent health expenditures (medicine, in/out-patient) in past month (KSh): **positive effect**

The 2016 paper (including 26 tables from its supplement), published in *Quarterly Journal of Economics*, doesn't report any of these.

	May 2011 draft	July 2011 draft	October 2011 draft	August 2012 draft	2016 paper
hemoglobin level	reported	reported	reported		
currently "very happy"	reported				
index of wellbeing	reported				
health expenditures	reported				

I emailed the authors and two of them replied, writing that it could be the referees, who didn't like too many variables being reported, the feedback from other economists, or it could simply that the paper became much different over the years with a lot of things being changed, not necessarily only these four variables. It could have been any of these things and **I do not say that the authors of this paper deliberately reported only the variables that favored their conclusion — they didn't. They reported a ton of variables, some supporting deworming, some not.**

However, two things to think about:

1. **The final 2016 paper looks as if these four variables were never collected in the first place.** A paper that reports 10 variables, 6 of them pointing in some direction and 4 not pointing in that direction is a very different paper, than the one that reports 6 variables, all pointing in that direction.
2. These are the four variables that were included in a public draft at some point and were later dropped. **How many variables were collected and were**

dropped before reaching any drafts?

My favorite quote from the final 2016 paper:

> In the full sample, treatment respondents' total nonagricultural earnings are 15.0% **higher** (112 shillings, std. err. 96, Table IV, Panel A), **although the effect is not statistically significant**. In the older than school age subsample, the effect is considerably larger at 22.6% (278 shillings, std. err. 167, **p = .101**). **[emphasis mine]**

Addendum: more misquotes and misrepresentations from *Doing Good Better*

These were either to difficult to explain in the main body or incidental and do not appear to be advancing the book's key claims.

Life satisfaction and income

MacAskill:

> In order to work out the relationship between level of income and level of subjective well-being, economists have conducted large-scale surveys of income levels and the subjective well-being of people in each of them. Their results are given in this graph, which shows the relationship between income and subjective well-being both within a country and across countries.

Graph

> The vertical axis of this graph represents self-reported well-being. Those interviewed had to say how satisfied they were with their lives on a scale from 0 to 10. Rating yourself at 10 means you consider yourself maximally happy: you think that, realistically, life couldn't get any better. Rating yourself at 0 means you consider yourself maximally unhappy: you think that, realistically, life couldn't get any worse. Most people fall in the middle of this range. The horizontal axis represents annual income.
>
> **What's interesting about this graph is that a doubling of income will always increase reported subjective well-being by the same amount.** For someone earning $1,000 per year, a $1,000 pay rise generates the same increase in happiness as a $2,000 pay rise for someone earning $2,000 per year, or an $80,000 pay rise for someone already earning $80,000 per year. And so on.
>
> **This graph allows us to determine just how much greater a benefit the extreme poor receive from one dollar than you or I do.** Imagine if your boss called you into her office and told you your salary would double for the next

> year. You'd be pretty pleased, right? What the conclusions from the economic studies [Note how a single paper morphed into "economic studies" — Alexey] suggest is that the benefit you get from having your salary doubled is the same as the benefit an extremely poor Indian farmer gets from having his salary doubled. If you're on the typical US wage of $28,000 per year, the benefit you'd get from an additional $28,000 in income is the same as the benefit a poor Indian farmer would get from an additional $220.
>
> This gives us a good theoretical reason for thinking that the same amount of money can do one hundred times as much to benefit the very poorest people in the world as it can to benefit typical citizens of the United States. **If you earn as much as the typical American worker, then you are one hundred times as rich as the very poorest people in the world, which means additional income can do a hundred times as much to benefit the extreme poor as it can to benefit you or me.** This isn't to say that income is all that matters to well-being—of course other factors like safety and political freedom are involved. But income certainly plays a critical role in how enjoyable, long, and healthy your life is. Looking at how much we can benefit people via increasing their income gives us a particularly robust way of assessing how much we can benefit others compared to ourselves.
>
> It's not often you have two options, one of which is one hundred times better than the other. Imagine a happy hour where you could either buy yourself a beer for five dollars or buy someone else a beer for five cents. If that were the case, we'd probably be pretty generous—next round's on me! But that's effectively the situation we're in all the time. It's like a 99-percent-off sale, or getting 10,000 percent extra free. It might be the most amazing deal you'll see in your life.
>
> **This idea is important enough that I've given it a name. I call it the 100x Multiplier.** For those of us living in rich countries, you should expect to be able to do **at least one hundred times** as much to benefit other people as you can to benefit yourself. **[emphasis mine]**

Here's the paper MacAskill cites.

Consider the following argument (Adapted from Westfall and Yarkoni, 2016):

> The sales of ice cream are positively correlated with temperature outside, thus to raise temperature outside we need to increase ice cream sales.

What's the difference between the argument above and the argument below?

> The income is positively correlated with happiness, thus to raise happiness we need to increase income.

The difference between the arguments is that the first is obviously false and the second is obviously true. That's the initial impression at least. In reality, the second argument seems right because its conclusion is intuitively right. When we evaluate it, we think "well, obviously income increases happiness!" and we completely forget about the premise and that the structure of the argument is "X is correlated with Y, therefore X causes Y". In fact, later in the book, *MacAskill makes the same exact point*:

> Of course, correlation is not causation. Merely showing that the people's welfare has improved at the same time the West has been offering aid does not prove that aid caused the improvement. It could be that aid is entirely incidental, or even harmful, holding back even greater progress that would have happened anyway or otherwise.

So MacAskill agrees that his argument about the causal relationship between life satisfaction and income is invalid. It seems that Stevenson and Wolfers agree as well, since **they never claimed to demonstrate the causal relationship with this graph alone.**

Therefore, I conclude that MacAskill knew that this argument is a misrepresentation of Stevenson and Wolfers but wrote it anyway.

Jobs that make a difference

MacAskill:

> Indeed, medicine is the banner career for people who want to make a difference. Every year, about twenty thousand people in the United States and eight thousand people in the United Kingdom go to medical school, and the number is growing year after year. **Even for those for whom medicine isn't a good fit, the desire to pursue a career that makes a difference is widespread. According to one study, 70 percent of young people regard ethical considerations as "crucial" in their choice of employer.** Enterprises like Teach for America have grown dramatically, explicitly targeting students who care more about making a difference than about making a high salary. Organizations like Net Impact, Idealist, and ethicalcareers.org all offer advice on choosing a vocation that does good. Even Oprah Winfrey, on her website, provides examples of "jobs that make a difference."

So, MacAskill uses the 70% number as evidence of people desiring to pursue a career that makes a difference. Now, look at the source article:

> Over 70% of students said that a company's ethical track record is a crucial factor when choosing their employer.

It has nothing to do with "making a difference".

Benefits from medicine in the US

(this seems to be just a mistake)

MacAskill:

> Greg found work by an epidemiologist named John Bunker, who estimated that the total benefits from medicine in the United States is about 7 QALYs per person

MacAskill cites "The Role of Medical Care in Contributing to Health Improvements within Societies," International Journal of Epidemiology 30, no. 6 (December 2001), 1,260–3.. The paper contains no such estimate.

Steve Jobs' interest in electronics

MacAskill writes:

> The evidence therefore suggests that following your passion is a poor way to determine whether a given career path will make you happy. Rather, passion grows out of work that has the right features. This was even true of Steve Jobs. **When he was young, he was passionate about Zen Buddhism. He traveled in India, took plenty of LSD, shaved his head, wore robes, and seriously considered moving to Japan to become a monk. He first got into electronics only reluctantly, as a way to earn cash on the side,** helping his tech-savvy friend Steve Wozniak handle business deals while also spending time at the All-One Farm. **[emphasis mine]**

For words *He traveled in India* MacAskill references Walter Isaacson, *Steve Jobs* (New York: Simon & Schuster, 2011), 39–50. These pages mention Jobs' travel (and other things), but never mention how he helped Steve Wozniak with business deals. They do however reference Jobs' previous interest in technology and electronics, directly contradicting MacAskill.

Page 43:

> In February 1974, after eighteen months of hanging around Reed, Jobs decided to move back to his parents' home in Los Altos and look for a job. It was not a difficult search. At peak times during the 1970s, the classified section of the San Jose Mercury carried up to sixty pages of technology help-wanted ads. One of those caught Jobs's eye. "Have fun, make money," it said. That day Jobs walked into the lobby of the video game manufacturer Atari and told the personnel director, who was startled by his unkempt hair and attire, that he wouldn't leave until they gave him a job. ...

> Jobs thus became one of the first fifty employees at Atari, working as a technician for $5 an hour. "In retrospect, it was weird to hire a dropout from Reed," Alcorn recalled. "But I saw something in him. **He was very intelligent, enthusiastic, excited about tech.**" **[emphasis mine]**

In February 1974, Jobs became a technician at Atari. This was two months before his India trip. No mention of Wozniak.

Page 48:

> [After Jobs' return from India] They [Jobs' parents] took him back home, where he continued trying to find himself. It was a pursuit with many paths toward enlightenment. **In the mornings and evenings he would meditate and study Zen, and in between he would drop in to audit physics or engineering courses at Stanford. [emphasis mine]**

This is late 1974. Jobs is auditing classes in physics and engineering at Stanford. No mention of Wozniak. It's more than 6 months until Wozniak's Apple I idea.

So, Steve Jobs was in fact interested in technology and electronics himself, *as the very pages that MacAskill cites show*.

Objection: Here, you could argue that it was just carelessness and it's wrong for me to call it a deliberate misrepresentation. MacAskill probably needed a reference to back up what he already had in the draft, so he (or his research assistant) just found a biography of Jobs and this is how Isaacson's book became referenced in *Doing Good Better*.

This interpretation is indefensible. First of all, It's not just the book that is referenced, but the specific pages. Furthermore, these pages (39-50) do not span a single complete chapter. Page 39 is closer to the end of the second chapter (The Dropout) while page 50 is in the middle of the third chapter (Atari and India), so the interpretation that MacAskill (or his research assistant) just found a chapter on India and LSD doesn't work either. **Somebody did look at the book, decided to reference these specific pages, and ignored the contradictions.**

Here are some more quotes from the two chapters I mentioned above:

Page 37:

> And even though he [Jobs] barely indulged it at Reed, **there was still an undercurrent of electronic geekiness in his soul** that would someday combine surprisingly well with the rest of the mix. **[emphasis mine]**

Page 55:

> Bushnell agreed. "There is something indefinable in an entrepreneur, and I saw that in Steve," he said. **"He was interested not just in engineering, but also the business aspects.** I taught him that if you act like you can do

> something, then it will work. I told him, 'Pretend to be completely in control and people will assume that you are.'" **[emphasis mine]**

Notice how close pages 37 and 55 are to referenced 39-50 and that they directly contradict MacAskill.

In addition to all of that, if we try to look at the rest of the book, we will find even more evidence that Jobs was interested in electronics for a very long time and didn't just want to make money off it. Few more examples:

Page 8:

> "The first computer terminal I ever saw was when my dad brought me to the Ames Center," he said. "I fell totally in love with it."

Page 10:

> Like most kids, he became infused with the passions of the grown ups around him. "Most of the dads in the neighborhood did really neat stuff, like photovoltaics and batteries and radar," Jobs recalled. "I grew up in awe of that stuff and asking people about it."

Page 16:

> He had few friends his own age, but he got to know some seniors who were immersed in the counterculture of the late 1960s. It was a time when the geek and hippie worlds were beginning to show some overlap. "My friends were the really smart kids," he said. **"I was interested in math and science and electronics.** They were too, and also into LSD and the whole counterculture trip." **[emphasis mine]**

Page 25:

> When it was finished, Fernandez told Wozniak there was someone at Homestead High he should meet. "His name is Steve. He likes to do pranks like you do, and he's also into building electronics like you are." It may have been the most significant meeting in a Silicon Valley garage since Hewlett went into Packard's thirty-two years earlier. "Steve and I just sat on the sidewalk in front of Bill's house for the longest time, just sharing stories—mostly about pranks we'd pulled, and also what kind of electronic designs we'd done," Wozniak recalled. "We had so much in common. Typically, it was really hard for me to explain to people what kind of design stuff I worked on, but Steve got it right away. And I liked him. He was kind of skinny and wiry and full of energy." Jobs was also impressed. "Woz was the first person I'd met who knew more electronics than I did," he once said, stretching his own expertise. "I liked him right away. I was a little more mature than my years,

and he was a little less mature than his, so it evened out. Woz was very bright, but emotionally he was my age."

Appendix

MacAskill:

> With the help of collaborators, Kremer tested the different ICS programs one by one. First, he looked at the efficacy of providing schools with additional textbooks. Classrooms would often have only one textbook for a class of thirty, so it seemed obvious that providing more textbooks would help students learn. However, when Kremer tested this theory by comparing test scores between schools that received books and those that didn't, he found no effect for all but the most high-achieving of students. (He suggests the textbooks were written at too high a level for the children, especially considering they were in English, the pupils' third language after Swahili and their local languages.)
>
> Next, Kremer looked at providing flip charts. The schoolchildren couldn't understand the textbooks, but having flip charts would allow teachers to tailor lessons to the specific needs of the students. Perhaps these would work better. Again, however, no effect.
>
> Undaunted, he took a different approach. If providing additional materials didn't work, maybe increasing the number of teachers would. After all, most schools had only one teacher, catering to a large class. But, again, he found no discernible improvement from decreasing class sizes.
>
> Over and over again, Kremer found that seemingly obvious programs to improve education just weren't working. But he persisted. He refused to believe there was simply no way to improve the education of children in Kenya. At that point, a friend at the World Bank suggested he test deworming.
>
> Few people in developed countries know about intestinal worms: parasitic infections that affect more than one billion people worldwide. They aren't as dramatic as AIDS or cancer or malaria, because they don't kill nearly as many people as those other conditions. But they do make children sick, and can be cured for pennies: off-patent drugs, developed in the 1950s, can be distributed through schools and administered by teachers, and will cure children of intestinal worms for a year.
>
> Kremer did an experiment to see whether treating children for these intestinal worms had an impact on education. The results were striking. "We didn't expect deworming to be as effective as it was," Kremer told me. "It

turned out to be one of the most cost-effective ways of increasing school participation."

Absenteeism is a chronic problem in schools in Kenya, and deworming reduced it by 25 percent. In fact, every child treated spent an extra two weeks in school, and every one hundred dollars spent on the program provided a total of ten years of additional school attendance among all students. Enabling a child to spend an extra day in school therefore cost just five cents. It wasn't merely that deworming children "worked" at getting children into school. It worked incredibly well.

What's more, deworming didn't merely have educational benefits. It had health and economic benefits, too. Intestinal worms can cause a variety of maladies, including anemia, intestinal obstruction, and a suppressed immune system that can increase the risk of other diseases like malaria. Deworming decreases all these risks.

Moreover, when Kremer's colleagues followed up with the children ten years later, those who had been dewormed were working an extra 3.4 hours per week and earning an extra 20 percent of income compared to those who had not been dewormed. In fact, deworming was such a powerful program that it paid for itself through increased tax revenue.

By the time his work on deworming was published, Kremer's revolutionary new approach to development had spawned a following, with dozens of the brightest young economists running hundreds of trials of different development programs. Meanwhile, Glennerster had quit her job and become the executive director of the newly founded Poverty Action Lab at MIT, where she used her knowledge of policy to ensure the research Kremer and his colleagues were conducting would have real-world impact.

In 2007, on the basis of this research, Kremer and Glennerster cofounded the nonprofit Deworm the World Initiative, which provides technical assistance to the governments of developing countries, enabling them to launch their own deworming programs. The charity has provided more than forty million deworming treatments, and the independent charity evaluator GiveWell regards them as one of the most cost-effective development charities.

-1

guzey

[-] **William_MacAskill** 36m 21

Hi Alexey,

I appreciate that you've taken the time to consider what I've said in the book at such length. However, I do think that there's quite a lot that's wrong in your post, and I'll describe some of that below. Though I think you have noticed a couple of mistakes in the book, I think that most of the alleged errors are not errors.

I'll just focus on what I take to be the main issues you highlight, and I won't address the 'dishonesty' allegations, as I anticipate it wouldn't be productive to do so; I'll leave that charge for others to assess.

tl;dr:

- Of the main issues you refer to, I think you've identified two mistakes in the book: I left out a caveat in my summary of the Baird et al (2016) paper, and I conflated overheads costs and CEO pay in a way that, on the latter aspect, was unfair to Charity Navigator.
- In neither case are these errors egregious in the way you suggest. I think that: (i) claiming that the Baird et al (2016) should cause us to believe that there is 'no effect' on wages is a misrepresentation of that paper; (ii) my core argument against Charity Navigator, regarding their focus on 'financial efficiency' metrics like overhead costs, is both successful and accurately depicts Charity Navigator.
- I don't think that the rest of the alleged major errors are errors. In particular: (i) GiveWell were able to review the manuscript before publication and were happy with how I presented their research; the quotes you give generally conflate how to think about GiveWell's estimates with how to think about DCP2's estimates; (ii) There are many lines of evidence supporting the 100x multiplier, and I don't rely at all on the DCP2 estimates, as you imply.

(Also, caveating up front: for reasons of time limitations, I'm going to have to precommit to this being my last comment on this thread.)

(Also, Alexey's post keeps changing, so if it looks like I'm responding to something that's no longer there, that's why.)

1. Deworming

Since the book came out, there has been much more debate about the efficacy of deworming. As I've continued to learn about the state and quality of the empirical evidence around deworming, I've become less happy with my presentation of the evidence around deworming in *Doing Good Better*; this fact has been reflected on the errata page on my website for the last two years. On your particular points, however:

Deworming vs textbooks

If textbooks have a positive effect, it's via how much children learn in school, rather than an incentive for them to spend more time in school. So the fact that there doesn't seem to be good evidence for textbooks increasing test scores is pretty bad.

If deworming has a positive effect, it could be via a number of mechanisms, including increased school attendance or via learning more in school, or direct health impacts, etc. If there are big gains on any of these dimensions, then deworming looks promising. I agree that more days in school certainly aren't good in themselves, however, so the better evidence is about the long-run effects.

Deworming's long-run effects

Here's how GiveWell describes the study on which I base my discussion of the long-run effects of deworming:

"10-year follow-up: Baird et al. 2016 compared the first two groups of schools to receive deworming (as treatment group) to the final group (as control); the treatment group was assigned 2.41 extra years of deworming on average. The study's headline effect is that as adults, those in the treatment group worked and earned substantially more, with increased earnings driven largely by a shift into the manufacturing sector." Then, later: "We have done a variety of analyses to assess the robustness

of the core findings from Baird et al. 2016, including reanalyzing the data and code underlying the study, and the results have held up to our scrutiny."

You are correct that my description of the findings of the Baird et al paper was not fully accurate. When I wrote, "Moreover, when Kremer's colleagues followed up with the children ten years later, those who had been dewormed were working an extra 3.4 hours per week and earning an extra 20 percent of income compared to those who had not been dewormed," I should have included the caveat "among non-students with wage employment." I'm sorry about that, and I'm updating my errata page to reflect this.

As for how much we should update on the basis of the Baird et al paper — that's a really big discussion, and I'm not going to be able to add anything above what GiveWell have already written (here, here and here). I'll just note that:

(i) Your gloss on the paper seems misleading to me. If you include people with zero earnings, of course it's going to be harder to get a statistically significant effect. And the data from those who do have an income but who aren't in wage employment are noisier, so it's harder to get a statistically significant effect there too. In particular, see here from the 2015 version of the paper: "The data on [non-agricultural] self-employment profits are likely measured with somewhat more noise. Monthly profits are 22% larger in the treatment group, but the difference is not significant (Table 4, Panel C), in part due to large standard errors created by a few male outliers reporting extremely high profits. In a version of the profit data that trims the top 5% of observations, the difference is 28% ($P < 0.10$)."

(ii) GiveWell finds the Baird et al paper to be an important part of the evidence behind their support of deworming. If you disagree with that, then you're engaged in a substantive disagreement with GiveWell's views; it seems wrong to me to class that as a simple misrepresentation.

2. Cost-effectiveness estimates

Given the previous debate that had occurred between us on how to think and talk about cost-effectiveness estimates, and the mistakes I had made in this regard, I wanted to be sure that I was presenting these estimates in a way that those at GiveWell would be happy with. So I asked an employee of GiveWell to look over the relevant parts of the manuscript of *DGB* before it was published; in the end five employees did so, and they were happy with how I presented GiveWell's views and research.

How can that fact be reconciled with the quotes you give in your blog post? It's because, in your discussion, you conflate two quite different issues: (i) how to represent that cost-effectiveness estimates provided by DCP2, or by single studies; (ii) how to represent the (in my view much more rigorous) cost-effectiveness estimates provided by GiveWell. Almost all the quotes from Holden that you give are about (i). But the quotes you criticise me for are about (ii). So, for example, when I say 'these estimates' are order of magnitude estimates that's referring to (i), not to (ii).

There's a really big difference between (i) and (ii). I acknowledge that back in 2010 I was badly wrong about the reliability of DCP2 and individual studies, and that GWWC was far too slow to update its web pages after the unreliability of these estimates came to light. But the level of time, care and rigour that have gone into the GiveWell estimates are much greater than those that have gone into the DCP2 estimates. It's still the case that there's a huge amount of uncertainty surrounding the GiveWell estimates, but describing them as "the most rigorous estimates" we have seems reasonable to me.

More broadly: Do I really think that you do as much good or more in expectation from donating $3500 to AMF as saving a child's life? Yes. GiveWell's estimate of the direct benefits might be optimistic or pessimistic (though it has stayed relatively stable over many years now — the median GiveWell estimate for 'cost for outcome as good as averting the death of an individual under 5' is currently $1932), but I really don't have a view on which is more likely. And, what's more important, the biggest consideration that's missing from GiveWell's analysis is the long-run effects of saving a life. While of course it's a thorny issue, I personally find it plausible that the long-run expected benefits from a donation to AMF are considerably larger than the short-run benefits — you speed up economic progress just a little bit, in expectation making those in the future just a little bit better off than they would have otherwise been. Because the future is so vast in expectation, that effect is very large. (There's *plenty* more to discuss on this issue of long-run effects — Might those effects be negative? How should you discount future consumption? etc — but that would take us too far afield.)

3. Charity Navigator

Let's distinguish: (i) the use of overhead ratio as a metric in assessing charities; (ii) the use of CEO pay as a metric in assessing charities. The idea of evaluating charities on overheads and on the basis of CEO pay are often run together in public discussion, and are both wrong for similar reasons, so I bundled them together in my discussion.

Regarding (ii): CN-of-2014 did talk a lot about CEO pay: they featured CEO pay, in both absolute terms and as a proportion of expenditure, prominently on their charity evaluation pages (see, e.g. their page on Books for Africa), they had top-ten lists like, "10 highly-rated charities with low paid CEOs", and "10 highly paid CEOs at low-rated charities" (and no lists of "10 highly-rated charities with high paid CEOs" or "10 low-rated charities with low paid CEOs"). However, it is true that CEO pay was not a part of CN's rating system. And, rereading the relevant passages of *DGB*, I can see how the reader would have come away with the wrong impression on that score. So I'm sorry about that. (Perhaps I was subconsciously still ornery from their spectacularly hostile hit piece on EA that came out while I was writing DGB, and was therefore less careful than I should have been.) I've updated my errata page to make that clear.

Regarding (i): CN's two key metrics for charities are (a) financial health and (b) accountability and transparency. (a) is in very significant part about the charities' overheads ratios (in several different forms), where they give a charity a higher score the lower its overheads are, breaking the scores into five broad buckets: see here for more detail. The doughnuts for police officers example shows that a really bad charity could score extremely highly on CN's metrics, which shows that CN's metrics must be wrong. Similarly for Books for Africa, which gets a near-perfect score from CN, and features in its 'ten top-notch charities' list, in significant part because of its very low overheads, despite having no good evidence to support its program.

I represent CN fairly, and make a fair criticism of its approach to assessing charities. In the extended quote you give, they caveat that very low overheads are not make-or-break for a charity. But, on their charity rating methodology, all other things being equal they give a charity a higher score the lower the charity's overheads. If that scoring method is a bad one, which it is, then my criticism is justified.

4. Life satisfaction and income and the hundredfold multiplier

The hundredfold multiplier

You make two objections to my 100x multiplier claim: that the DCP2 deworming estimate was off by 100x, and that the Stevenson and Wolfers paper does not support it.

But there are very many lines of evidence in favour of the 100x multiplier, which I reference in *Doing Good Better*. I mention that there are many independent justifications for thinking that there is a logarithmic (or even more concave) relationship between income and happiness on p.25, and in the endnotes on p.261-2 (all references are to the British paperback edition - yellow cover). In addition to the Stevenson and Wolfers lifetime satisfaction approach (which I discuss later), here are some reasons for thinking that the hundredfold multiplier obtains:

- The experiential sampling method of assessing happiness. I mention this in the endnote on p.262, pointing out that, on this method, my argument would be stronger, because on this method the relationship between income and wellbeing is more concave than logarithmic, and is in fact bounded above.
- Imputed utility functions from the market behaviour of private individuals and the actions of government. It's absolutely mainstream economic thought that utility varies with log of income (that is, eta=1 in an isoelastic utility function) or something more concave (eta>1). I reference a paper that takes this approach on p.261, the Groom and Maddison (2013). They estimate eta to be 1.5.
- Estimates of cost to save a life. I discuss this in ch.2; I note that this is another strand of supporting evidence prior to my discussion of Stevenson and Wolfers on p.25: "It's a basic rule of economics that money is less valuable to you the more you have of it. We should therefore expect $1 to provide a larger benefit for an extremely poor Indian farmer than it would for you or me. But how much larger? Economists have sought to answer this question through a variety of methods. We'll look at some of these in the next chapter, but for now I'll just discuss one [the Stevenson and Wolfers approach]." Again, you find 100x or more discrepancy in the cost to save a life in rich or poor countries.
- Estimate of cost to provide one QALY. As with the previous bullet point.

Note, crucially, that the developing world estimates for cost to provide one QALY or cost to save a life come from GiveWell, **not** — as you imply — from DCP2 or any individual study.

Is there a causal relationship from income to wellbeing?

It's true that there Stevenson and Wolfers only shows the correlation is between income and wellbeing. But that there is a causal relationship, from income to wellbeing, is beyond doubt. It's perfectly obvious that, over the scales we're talking, higher income enables you to have more wellbeing (you can buy analgesics, healthcare, shelter, eat more and better food, etc).

It's true that we don't know exactly the strength of the causal relationship. Understanding this could make my argument stronger or weaker. To illustrate, here's a quote from another Stevenson and Wolfers paper, with the numerals in square brackets added in by me:

"Although our analysis provides a useful measurement of the bivariate relationship between income and well-being both within and between countries, there are good reasons to doubt that this corresponds to the causal effect of income on well-being. It seems plausible (perhaps even likely) that [i] the within-country well-being-income gradient may be biased upward by reverse causation, as happiness may well be a productive trait in some occupations, raising income. A different perspective, from offered by Kahneman, et al. (2006), suggests that [ii] within-country comparisons overstate the true relationship between subjective well-being and income because of a "focusing illusion": the very nature of asking about life satisfaction leads people to assess their life relative to others, and they thus focus on where they fall relative to others in regard to concrete measures such as income. Although these specific biases may have a more important impact on within-country comparisons, it seems likely that [iii] the bivariate well-being-GDP relationship may also reflect the influence of third factors, such as democracy, the quality of national laws or government, health, or even favorable weather conditions, and many of these factors raise both GDP per capita and well-being (Kenny, 1999).29 [iv] Other factors, such as increased savings, reduced leisure, or even increasingly materialist values may raise GDP per capita at the expense of subjective well-being. At this stage we cannot address these shortcomings in any detail, although, given our reassessment of the stylized facts, we would suggest an urgent need for research identifying these causal parameters."

To the extent to which (i), (ii) or (iv) are true, the case for the 100x multiplier becomes stronger. To the extent to which (iii) is true, the case for the 100x multiplier becomes weaker. We don't know, at the moment, which of these are the most important factors. But, given that the wide variety of different strands of evidence listed in the previous section all point in the same direction, I think that estimating a 100x multiplier as a causal matter is reasonable. (Final point: noting again that all these estimates do not factor in the long-run benefits of donations, which would increase the ratio of benefits others to benefits to yourself even further in the direction of benefits to others.)

On the Stevenson and Wolfers data, is the relationship between income and happiness weaker for poor countries than for rich countries?

If it were the case that money does less to buy happiness (for any given income level) in poor countries than in rich countries, then that would be one counterargument to mine.

However, it doesn't seem to me that this is true of the Stevenson and Wolfers data. In particular, it's highly cherry-picked to compare Nigeria and the USA as you do, because Nigeria is a clear outlier in terms of how flat the slope is. I'm only eyeballing the graph, but it seems to me that, of the poorest countries represented (PHL, BGD, EGY, CHN, IND, PAK, NGA, ZAF, IDN), only NGA and ZAF have flatter slopes than USA (and even for ZAF, that's only true for incomes less than $6000 or so); all the rest have slopes that are similar to or steeper than that of USA (IND, PAK, BGD, CHN, EGY, IDN all seem steeper than USA to me). Given that Nigeria is such an outlier, I'm inclined not to give it too much weight. The average trend across countries, rich and poor, is pretty clear.

[-] **guzey** 5m 1

Hi William,

Thank you for your response. I apologize for the stronger language that I used in the first public version of this post. I believe that here you do not address most of the points I made either in the first public version or in the version that was up here at the moment of your comment.

I will not change the post here without explicitly noting it, now that you have replied.

I'm in the process of preparing a longer reply to you.

> [-] **guzey** 2m 1
>
> In particular, the version of the essay that I initially posted here did not discuss the strength of the relationship between income and happiness in rich and poor countries -- I agree that this was a weak argument.

[-] **SiebeRozendal** 2h 3

I admire the amount of effort that has gone into this post and its level of rigor. I think it's very important for an epistemically healthy movement that high-status people can be criticised successfully.

I think your premises do not fully support the conclusion that MacAskill is completely untrustworthy. However, I agree that the book misrepresents sources structurally, and this is a convincing sign it is written in bad faith.

I hope that MacAskill has already realized the book was not up to the standards he now promotes. Writing an introduction to effective altruism was and remains a very difficult task, and at the time there was still a mindset of "push EA even if it's at the cost of some epistemic honesty". I think the community has been moving away from this mindset since, and this post is a good addition to that.

We need a better introductory book. (Also because it's outdated.)

> [-] **guzey** 2h 2
>
> Thanks. I agree with you that it does not show complete untrustworthiness. Adjusted the language a little bit.

> [-] **Denise_Melchin** 4m 2
>
> I don't think unsuccessful applications at organizations that are distantly related to the content you're criticizing constitute a conflict of interest.
>
> If everybody listed their unsuccessful applications at the start of every EA Forum post, it would take up a lot of reader attention.
>
> > [-] **guzey** now 1
> >
> > I heavily criticize one of the founders of CEA and heavily use the words of the founder of Open Phil in my post, which lead me to believe that I need to disclose that I applied to both organizations.

[+] **rafa_fanboy** 4h -17