

# The Institutional Critique of Effective Altruism

BRIAN BERKEY

*University of Pennsylvania*

In recent years, the effective altruism movement has generated much discussion about the ways in which we can most effectively improve the lives of the global poor, and pursue other morally important goals. One of the most common criticisms of the movement is that it has unjustifiably neglected issues related to institutional change that could address the root causes of poverty, and instead focused its attention on encouraging individuals to direct resources to organizations that directly aid people living in poverty. In this article, I discuss and assess this 'institutional critique'. I argue that if we understand the core commitments of effective altruism in a way that is suggested by much of the work of its proponents, and also independently plausible, there is no way to understand the institutional critique such that it represents a view that is both independently plausible and inconsistent with the core commitments of effective altruism.

In recent years, the effective altruism movement has generated a great deal of discussion about the ways in which we (that is, at least those of us who are at least reasonably well off) can most effectively improve the lives of the global poor, and pursue other morally important goals (for example, reducing the suffering of non-human animals, or achieving criminal justice reform in the United States).<sup>1</sup> Two of the movement's most prominent members, Peter Singer and William MacAskill, have each produced a popular book aimed at broad audiences.<sup>2</sup> These books, and the efforts of the movement that they represent, have in turn generated a significant amount of critical discussion. This discussion has, to this point, taken place primarily in popular media outlets and on blogs, rather than in academic journals.<sup>3</sup>

<sup>1</sup> On criminal justice reform, see William MacAskill, *Doing Good Better: How Effective Altruism Can Help You Make a Difference* (New York, 2015), pp. 185–7. For information on GiveWell's work in this area, see <<http://www.givewell.org/labs/causes/criminal-justice-reform>>.

<sup>2</sup> Peter Singer, *The Most Good You Can Do: How Effective Altruism is Changing Ideas about Living Ethically* (New Haven, 2015); MacAskill, *Doing Good Better*.

<sup>3</sup> Important exceptions are Theron Pummer, 'Whether and Where to Give', *Philosophy and Public Affairs* 44 (2016), pp. 77–95, and Iason Gabriel, 'Effective Altruism and its Critics', *Journal of Applied Philosophy* (forthcoming). In addition, the *Journal of Global Ethics* recently published a symposium on Singer's *The Most Good You Can Do* (vol. 12, no. 2), and a recent special issue of *Essays in Philosophy* was dedicated to Effective Altruism (vol. 18, no. 1). Another valuable contribution is Jennifer Rubenstein, 'The Lessons of Effective Altruism', *Ethics & International Affairs* 30 (2016), pp. 511–26, which is an extended review of Singer and MacAskill's books.

One of the most common criticisms of the movement is that it has unjustifiably neglected issues related to institutional change that could address the root causes of poverty, and chosen instead to focus its attention on encouraging individuals to direct resources to organizations that directly aid people living in poverty around the world (by providing them with, for example, bednets that protect them from malaria,<sup>4</sup> or deworming treatments,<sup>5</sup> or direct cash transfers<sup>6</sup>). This complaint has been made in somewhat different ways by different critics,<sup>7</sup> but the unifying thought seems to be that there is something about institutional change that is distinctively morally important in

<sup>4</sup> See GiveWell's recommendation of the Against Malaria Foundation: <<http://www.givewell.org/international/top-charities/amf>>.

<sup>5</sup> See GiveWell's recommendation of the Deworm the World Initiative: <<http://www.givewell.org/international/top-charities/deworm-world-initiative>>.

<sup>6</sup> See GiveWell's recommendation of GiveDirectly: <<http://www.givewell.org/International/top-charities/give-directly>>.

<sup>7</sup> See, for example, the contributions to the *Boston Review's* July 2015 forum on 'The Logic of Effective Altruism' by Daron Acemoglu (<<https://bostonreview.net/forum/logic-effective-altruism/daron-acemoglu-response-effective-altruism>>), Angus Deaton (<<https://bostonreview.net/forum/logic-effective-altruism/angus-deaton-response-effective-altruism>>), Iason Gabriel (<<https://bostonreview.net/forum/logic-effective-altruism/iason-gabriel-response-effective-altruism>>) and Catherine Tumber (<<https://bostonreview.net/forum/logic-effective-altruism/catherine-tumber-response-effective-altruism>>); Emily Clough, 'Effective Altruism's Political Blind Spot', *Boston Review* (July 2015), <<<https://bostonreview.net/world/emily-clough-effective-altruism-ngos>>>; Pete Mills's contribution to the *Oxford Left Review's* forum on 'The Ethical Careers Debate' (no. 7, May 2012): <<https://oxfordleftreview.files.wordpress.com/2012/07/issue-7-new.pdf>>; Amia Srinivasan, 'Stop the Robot Apocalypse', *London Review of Books* 37 (2015), pp. 3–6 (<<http://www.lrb.co.uk/v37/n18/amia-srinivasan/stop-the-robot-apocalypse>>); Judith Lichtenberg, 'Peter Singer's Extremely Altruistic Heirs', *New Republic* (November 2015): <<https://newrepublic.com/article/124690/peter-singers-extremely-altruistic-heirs>>; Robert Mark Simpson, 'Moral Renegades', *The New Rambler* (July 2016): <<http://newramblerreview.com/book-reviews/philosophy/moral-renegades>>; Lisa Herzog, 'Can 'Effective Altruism' Really Change the World?', *OpenDemocracy.net* (February 2016): <<https://www.opendemocracy.net/transformation/lisa-herzog/can-effective-altruism-really-change-world#>> and '(One of) Effective Altruism's Blind Spot(s), or: Why Moral Theory Needs Institutional Theory', *Justice Everywhere Blog* (October 2015): <<http://justice-everywhere.org/international/one-of-effective-altruisms-blind-spots-or-why-moral-theory-needs-institutional-theory/>>; Grace Boey, 'Effective Altruism and its Blind Spots', 3 Quarks Daily (August 2015): <<http://www.3quarksdaily.com/3quarksdaily/2015/08/effective-altruism-and-its-blind-spots.html>>; Sam Earle and Rupert Read, 'Why Effective Altruism is Ineffective: The Case of Refugees', *Ecologist.org* (April 2016): <[http://www.theecologist.org/essays/2987460/why\\_effective\\_altruism\\_is\\_ineffective\\_the\\_case\\_of\\_refugees.html](http://www.theecologist.org/essays/2987460/why_effective_altruism_is_ineffective_the_case_of_refugees.html)>; Pascal-Emmanuel Gobry, 'Can Effective Altruism Really Change the World', *The Week* (March 2015): <<http://theweek.com/articles/542955/effective-altruism-really-change-world>>; Harriet Lamb, 'Effective Altruism is Good – Changing the System is Better', *Huffington Post* (May 2015): <[http://www.huffingtonpost.co.uk/harriet-lamb/effective-altruism-is-good\\_b\\_7933282.html](http://www.huffingtonpost.co.uk/harriet-lamb/effective-altruism-is-good_b_7933282.html)>; for a response from a proponent of effective altruism, see Robert Wiblin, 'Effective Altruists Love Systemic Change', *80,000 Hours Blog* (July 2015): <<https://80000hours.org/2015/07/effective-altruists-love-systemic-change/>>; for general discussion, see Gabriel,

the face of the facts on the ground regarding global poverty, and that this is something that effective altruists have tended to overlook, at least in part due to their acceptance of certain moral and/or empirical claims that ought to be rejected.

My aim in this article is to discuss and assess this ‘institutional critique’ of effective altruism. In order to do this, it will be necessary to consider exactly how we should understand the critique, and how it might challenge one or more of the core commitments of effective altruism. I will argue that if we understand the core commitments of effective altruism in a way that is suggested by much of the work of its proponents, and also independently plausible, there is no way to understand the institutional critique such that it represents a view that is both independently plausible and inconsistent with at least one of the core commitments of effective altruism.<sup>8</sup>

The article will proceed as follows. First, in [section I](#), I will describe and explain what seem to me to be the core commitments of effective altruism, drawing on work by prominent members of the movement, and focusing on issues related to global poverty.<sup>9</sup> Next, in [section II](#), I will examine the ways in which various critics have articulated versions of the institutional critique, and consider how that critique might be best formulated and defended. In [section III](#), I will argue that any version of the institutional critique that represents a view that is in fact inconsistent with at least one of the core commitments of effective altruism will have unacceptable implications regarding the range of moral reasons that we can have for acting to improve the lives of the global poor, and should therefore be rejected. I will conclude, in [section IV](#), by suggesting some reasons why some may find the

‘Effective Altruism and its Critics’. An early version of the institutional critique, which focuses on efforts to alleviate hunger, and is aimed at arguments made by Singer (‘Famine, Affluence, and Morality’, *Philosophy and Public Affairs* 1 (1972), pp. 229–43); Peter Unger (*Living High and Letting Die: Our Illusion of Innocence* (New York, 1996)); and Garrett Cullity (‘International Aid and the Scope of Kindness’, *Ethics* 105 (1994), pp. 99–127), can be found in Paul Gomberg, ‘The Fallacy of Philanthropy’, *Canadian Journal of Philosophy* 32 (2002), pp. 29–66.

<sup>8</sup> Despite rejecting what I call the ‘institutional critique’, I do not claim that effective altruists should not focus their efforts primarily on institutional reform. Indeed, it is consistent with my argument that the core commitments of effective altruism support such a focus. My argument does not, then, commit me to denying that the Open Philanthropy Project’s work on various US policy issues (see <<http://www.openphilanthropy.org/focus/us-policy>>) counts as effective altruism, or that it might be justified by the core commitments of effective altruism.

<sup>9</sup> As I noted above, effective altruists work to address a range of morally important issues in addition to global poverty. I focus my analysis on global poverty primarily because it is the issue most commonly discussed by those who advance versions of the institutional critique, but also for ease of exposition.

institutional critique appealing, despite its flaws, and explaining why this appeal is illusory.

## I. THE CORE COMMITMENTS OF EFFECTIVE ALTRUISM

MacAskill says that effective altruism

is about asking, ‘How can I make the biggest difference I can?’ and using evidence and careful reasoning to try to find an answer. It . . . consists of the honest and impartial attempt to work out what’s best for the world, and a commitment to do what’s best, whatever that turns out to be.<sup>10</sup>

Singer refers to a definition according to which effective altruism is ‘a philosophy and social movement which applies evidence and reason to working out the most effective ways to improve the world’, and says that effective altruists would all agree that ‘a world with less suffering and more happiness in it is, other things being equal, better than one with more suffering and less happiness’.<sup>11</sup> Robert Wiblin, who works at the effective altruist organization 80,000 Hours, describes effective altruism as ‘the use of evidence and analysis to take actions that help others as much as possible’.<sup>12</sup> These descriptions suggest that what it is to be an effective altruist is to take there to be strong reasons to act to improve the world, in particular by reducing suffering and increasing happiness, as well as reasons to prefer producing greater improvements over smaller ones, and reasons to gather and deploy relevant evidence in order to determine, as best one can, which of the actions available to one is in fact likely to produce the greatest improvements. It seems to me, then, that the core commitments of effective altruism can plausibly be represented in the following four claims:

*EA1:* There are very strong moral reasons, grounded in fundamental values, for the well off to direct significant resources to efforts to address important moral issues (e.g. to alleviate the plight of the global poor).

*EA2:* These fundamental values include (but are not necessarily limited to)<sup>13</sup> impartially promoting increases in welfare, or quality of life, for individuals, and the reasons provided by this value are at least fairly weighty.

<sup>10</sup> MacAskill, *Doing Good Better*, p. 11.

<sup>11</sup> Singer, *The Most Good You Can Do*, pp. 4–7.

<sup>12</sup> Robert Wiblin, ‘Disagreeing about What’s Effective Isn’t Disagreeing With Effective Altruism’, *80,000 Hours Blog* (July 2015): <<https://80000hours.org/2015/07/disagreeing-about-whats-effective-isnt-disagreeing-with-effective-altruism/>>.

<sup>13</sup> See Singer, *The Most Good You Can Do*, pp. 8–9.

*EA3:* There are strong reasons to prefer giving to efforts that will promote the relevant values most efficiently.

*EA4:* We should employ the best empirical research methods available in order to determine, as best we can, which efforts promote those values most efficiently.

It is easy to see why those who endorse these core commitments would recommend contributing to the kinds of efforts that effective altruists tend to promote. There is strong evidence, for example, that distributing insecticide-treated bednets in malaria zones is a cost-effective way of reducing the number of people who become infected with malaria, and thereby reducing child mortality in the areas in which the nets are distributed.<sup>14</sup> Since both living with malaria and dying from it significantly affect the welfare of the impoverished people who get it for the worse, effective altruists typically hold that there are strong moral reasons to donate money to the Against Malaria Foundation, which has a strong record of transparency and success in distributing nets to their intended recipients. Similarly, there seems to be good reason to believe that providing direct cash transfers to impoverished people is a cost-effective way of improving their lives.<sup>15</sup> Since GiveDirectly has a strong record of effectively transferring cash to extremely low-income households using mobile-phone based payment services, many effective altruists hold that there are strong moral reasons to donate money to that organization.<sup>16</sup>

It is also easy to see why those who endorse EA1–EA4 would be critical of, for example, giving money to wealthy universities,<sup>17</sup> or to organizations that train guide dogs for the blind.<sup>18</sup> Donations to wealthy universities will, generally, simply increase their already large

<sup>14</sup> MacAskill, *Doing Good Better*, pp. 53, 81–2; GiveWell recommendation of Against Malaria Foundation (<<http://www.givewell.org/international/top-charities/amf>>).

<sup>15</sup> GiveWell recommendation of GiveDirectly (<<http://www.givewell.org/International/top-charities/give-directly>>).

<sup>16</sup> MacAskill claims that the overall balance of the available evidence suggests that there are more efficient ways of improving the lives of impoverished people than donating to GiveDirectly, but acknowledges that this conclusion can be reasonably disputed (*Doing Good Better*, p. 120). Unlike GiveWell, the effective altruist organization Giving What We Can (<<https://www.givingwhatwecan.org/>>) does not list GiveDirectly among its top charities (<<https://www.givingwhatwecan.org/top-charities/>>). Their concerns are expressed on their website in a piece by MacAskill (<<https://www.givingwhatwecan.org/blog/2012-11-30/givewell%E2%80%99s-recommendation-of-givedirectly>>), and more recently in another by Andreas Mogensen (<<https://www.givingwhatwecan.org/blog/2014-02-27/why-we-still-don%E2%80%99t-recommend-givedirectly>>).

<sup>17</sup> Singer, *The Most Good You Can Do*, pp. 10–11.

<sup>18</sup> Singer, 'The Why and How of Effective Altruism', TED Talk (March 2013): <[http://www.ted.com/talks/peter\\_singer\\_the\\_why\\_and\\_how\\_of\\_effective\\_altruism/transcript?language=en](http://www.ted.com/talks/peter_singer_the_why_and_how_of_effective_altruism/transcript?language=en)>.

endowments. If doing this has any positive effects at all, it seems very likely that such effects will be extremely small in comparison to the benefits that could be obtained by directing the same funds to an organization recommended by effective altruists. And while training a guide dog that will improve the life of one blind person is surely a good thing to do, the cost of doing so is approximately \$40,000. At the same time, the cost of curing one person with trachoma in the developing world of blindness is approximately \$20–50. This means that for the same amount of money, one can either provide one guide dog to one blind person in the developed world, or cure between 800 and 2,000 people of blindness in the developing world. Since each dollar donated to curing blindness in the developing world will clearly improve the quality of life of those aided much more efficiently than each dollar donated to training guide dogs, those who accept EA2 and EA3 will strongly favour donating to efforts to cure blindness in the developing world.

## II. THE INSTITUTIONAL CRITIQUE

In response to effective altruists, at least most of whom endorse the view that individuals have strong reasons to donate significant amounts of money to the Against Malaria Foundation, GiveDirectly, and a small number of other highly effective organizations, a number of critics have argued that the focus of their efforts to address the plight of the global poor is largely misplaced. These critics claim that those who are motivated to contribute to addressing the problems of global poverty should direct their efforts at least primarily towards institutional reform, rather than to organizations that provide direct aid to those living in poverty, such as many of those recommended by effective altruists.

Consider the following ways in which versions of this ‘institutional critique’ have been articulated by critics of effective altruism. Judith Lichtenberg claims that

[T]he maximum effectiveness strategy [endorsed by effective altruists] means neglecting programs that support advocacy for political and structural change, which are essential for addressing the deeper roots of poverty . . . People across the political spectrum should agree that structural changes that allow all workers to earn a decent living are preferable to welfare programs and private charity.<sup>19</sup>

<sup>19</sup> Lichtenberg, ‘Peter Singer’s Extremely Altruistic Heirs’.

Lisa Herzog claims that effective altruists have a ‘picture [of] the social world and human institutions’ that is ‘flawed’.<sup>20</sup> She interprets them as taking the social world to consist of

single individuals, with independently formed utility functions, [who] choose between different options in ways that maximize their utility . . . basically ‘economic man’, the figure used in economic modeling – except that in most economic models, utility functions contain only one’s own utility, whereas here they contain the wellbeing of other people or animals.<sup>21</sup>

In her view, this picture is seriously misleading, since ‘human beings are far less rational . . . and how [they] behave depends on the social settings [they] find [themselves] in . . . They live in social structures that they inherit from the past . . . [These structures] are undergirded by institutions’.<sup>22</sup> She takes this alternative, non-atomistic picture of the social world to support the view that ‘[w]hat matters for a good human life, in which basic needs can be justified and individuals have some degree of autonomy, is that these institutions and practices function to the advantage of everyone’.<sup>23</sup> Because of this, she thinks that ‘the most dangerous underlying assumption in the worldview of effective altruists [is] that they take the current institutional order as given, implicitly denying that it is open to change’.<sup>24</sup> This assumption is seriously problematic, she claims, because ‘for change to be effective, we need to change the institutions and practices of today’s world . . . one of our greatest responsibilities is to try to change the structures of capitalism-gone-wild that does so much harm’.<sup>25</sup>

Others have also suggested that effective altruists have tended to be unduly complacent towards the prevailing institutions of global capitalism, or more strongly, have in fact allied themselves with those institutions, despite their role in perpetuating the suffering that they aim to relieve. In his criticism of 80,000 Hours’s claim that individuals should consider choosing a career in which they will earn a very large salary, such as investment banking, corporate law, or management consulting, so as to be able to donate a significant portion of that salary to effective aid organizations,<sup>26</sup> Pete Mills argues that:

professional philanthropy [that is, earning to give] does not just involve making your peace with the system – it means embracing it. The unstated imperative:

<sup>20</sup> Herzog, (One of) Effective Altruism’s Blind Spot(s).

<sup>21</sup> Herzog, (One of) Effective Altruism’s Blind Spot(s).

<sup>22</sup> Herzog, (One of) Effective Altruism’s Blind Spot(s).

<sup>23</sup> Herzog, (One of) Effective Altruism’s Blind Spot(s).

<sup>24</sup> Herzog, (One of) Effective Altruism’s Blind Spot(s).

<sup>25</sup> Herzog, (One of) Effective Altruism’s Blind Spot(s).

<sup>26</sup> <<https://80000hours.org/articles/earning-to-give/>>.



don't rock the boat . . . what enriches you is your position in a set of profoundly exploitative social relations, which we might label capitalism.<sup>27</sup>

He claims that because those who earn to give rely for their ability to give significant amounts of money to effective aid organizations on their privileged position within an unjust global economic order, '[t]he good that a professional philanthropist does depends on perpetuating a system which immiserates a vast portion of the world's population. The result is a toxic political quietism . . . In practice, 80k demands the systematic foreclosure of political alternatives'.<sup>28</sup> Echoing elements of both Herzog and Lichtenberg's critiques, he adds that

[w]ithout any concept of society as a collective endeavor, we cannot address problems at their root but only those symptoms which are tractable on an atomized, individual level . . . Poverty is presented to us as an immediate ethical demand which obscures the need for systemic change.<sup>29</sup>

Finally, consider how Amia Srinivasan, in her review of MacAskill's book, attempts to tie effective altruists' commitment to relying on the best available empirical evidence, and rough attempts to estimate the amount of good that can be done by directing resources to different kinds of efforts and organizations, to what she views as a problematic complacency regarding the institutions of global capitalism. She acknowledges that effective altruists have considered whether contributing to efforts to promote institutional change might be a strategy for improving the world the potential effectiveness of which the available evidence could support.<sup>30</sup> But she suggests that it is troubling that, for effective altruists, '[e]ven in these cases, the numbers are what matter'.<sup>31</sup> She notes that MacAskill claims that given the odds of an Oxford PPE student succeeding in an attempt to win elected office, and the amount of money the allocation of which she would have some influence over, going into politics appears to be a reasonable choice for such a student from the perspective of effective altruism. But, she claims, the same could not be said by an effective altruist in defence of, for example, contributing to efforts to promote radical political change:

<sup>27</sup> Mills, 'The Ethical Careers Debate', p. 5.

<sup>28</sup> Mills, 'The Ethical Careers Debate', p. 5.

<sup>29</sup> Mills, 'The Ethical Careers Debate', p. 8.

<sup>30</sup> She notes that some effective altruists have taken up immigration reform and the reform of laws governing factory farming practices as areas of focus in their efforts to improve the world ('Stop the Robot Apocalypse'). As I noted above, criminal justice reform in the United States is another issue on which some effective altruists have focused their attention.

<sup>31</sup> Srinivasan, 'Stop the Robot Apocalypse'.



What's the expected marginal value of becoming an anti-capitalist revolutionary? To answer that you'd need to put a value and probability measure on achieving an unrecognizably different world – even, perhaps, on our becoming unrecognizably different sorts of people. It's hard enough to quantify the value of a philanthropic intervention: how would we go about quantifying the consequences of radically reorganizing society?<sup>32</sup>

More generally, she suggests that the effective altruists' approach to thinking about how best to improve the world reflects, and amounts to an implicit embrace of, the global capitalist system that is ultimately responsible for the suffering that they hope to mitigate:

MacAskill is evidently comfortable with ways of talking that are familiar from the exponents of global capitalism: the will to quantify, the essential comparability of all goods and evils, the obsession with productivity and efficiency . . . There is a seemingly unanswerable logic, at once natural and magical, simple and totalizing, to both global capitalism and effective altruism. That he speaks in the proprietary language of the illness – global inequality – whose symptoms he proposes to mop up is an irony on which he doesn't comment.<sup>33</sup>

Commenting on his possible motivations for neglecting to direct a greater amount of critical attention to the system whose bad effects he hopes to mitigate, Srinivasan suggests that '[p]erhaps [MacAskill] senses that his potential followers – privileged, ambitious millennials – don't want to hear about the iniquities of the system that has shaped their worldview'.<sup>34</sup> She claims that because it has not focused sufficient attention on institutional issues, and in particular on the ways in which global capitalism has generated the problems that it hopes to help address, 'effective altruism, so far at least, has been a conservative movement, calling us back to where we already are: the world as it is, our institutions as they are'.<sup>35</sup> She adds that:

MacAskill does not address the deep sources of global misery . . . or the forces that ensure its reproduction. Effective altruism doesn't try to understand how power works, except to better align itself with it. In this sense it leaves everything just as it is. This is no doubt comforting to those who enjoy the status quo – and may in part account for the movement's success.<sup>36</sup>

There are a number of common threads that can be identified in the criticisms that I have noted. Herzog and Mills both argue that effective altruists operate with a picture of individuals as isolated, atomized agents, rather than as interconnected parts of a social world capable of

<sup>32</sup> Srinivasan, 'Stop the Robot Apocalypse'.

<sup>33</sup> Srinivasan, 'Stop the Robot Apocalypse'.

<sup>34</sup> Srinivasan, 'Stop the Robot Apocalypse'.

<sup>35</sup> Srinivasan, 'Stop the Robot Apocalypse'.

<sup>36</sup> Srinivasan, 'Stop the Robot Apocalypse'.

working together to change it. Srinivasan, like Mills, seems to believe that in recommending at least some of the courses of action that they do, effective altruists are, in effect, endorsing the prevailing global capitalist institutional order. This is problematic, they argue, because it is precisely that institutional order that must be changed if we are to address the problems of global poverty in a meaningful way. Herzog and Srinivasan both criticize effective altruists for suggesting that we ought to attempt, as best we can, to estimate the expected value of alternative approaches that an individual might take to trying to improve the world. And Lichtenberg, Herzog, Mills and Srinivasan all claim that effective altruists, in effect, take the existing global institutional order as given, and perform their analyses of what individuals can do to have the largest positive effect on the world with the continued existence of that institutional order as a background assumption. This, they seem to think, problematically limits the range of options that might be considered 'effective' by the analyses.

There are two core claims that all of the proponents of the institutional critique seem to endorse:

*IC1:* There are strong moral reasons for individuals to direct resources and time to efforts to promote institutional reform, rather than directing the same resources and time to providing aid to those living in poverty (e.g. by donating to some of the organizations recommended by effective altruists).

*IC2:* Effective altruists cannot, given their core commitments, support individuals directing resources and/or time to at least some of the efforts to promote institutional change to which there are in fact good moral reasons to devote resources/time.<sup>37</sup>

The success of the institutional critique requires that both of these two claims are true.

I will argue that, in order for both of the claims to be true, there would need to be good reason to reject at least one of the core commitments of effective altruism that I described above (that is, EA1–EA4). Proponents of the institutional critique are, I will claim, most plausibly interpreted as committed to rejecting either or both

<sup>37</sup> A closely related claim that at least some of the critics might have in mind is that even if the core commitments of effective altruism do not necessarily preclude supporting individuals devoting resources to efforts to promote institutional reform that there are good moral reasons to support (so that effective altruists might, in principle, be able to support them for reasons that are internal to their moral outlook), those core commitments do preclude endorsing, or giving sufficient weight to, certain important moral values that are among the appropriate *grounds* for supporting the relevant efforts to promote institutional reform.

of EA2 and EA4. I will argue, however, that any view that rejects either of these commitments is unacceptable, and that therefore the institutional critique fails.

### III. ASSESSING THE INSTITUTIONAL CRITIQUE

Herzog and Srinivasan appear to think that the tendency of effective altruists to frame their analyses of what individuals might do to improve the world the most in expected value terms biases those analyses against efforts to promote institutional reform. For Herzog, the source of the problem lies in effective altruists' allegedly atomistic picture of the social world, whereas for Srinivasan, it lies in their insistence on attempting to quantify the value of actions in terms of their potential effects and probabilities of success. Herzog thinks that by viewing individuals as isolated atoms who must each decide how they can best improve the world, holding fixed how they can expect others to behave, effective altruists commit themselves to viewing existing institutional arrangements as, in effect, unchangeable, so that the only options left for improving the world involve doing things like 'spend[ing] some of our spare money to help repair the worst damage that this system does in other parts of the world'.<sup>38</sup> And Srinivasan suggests that the effective altruist commitment to estimating probabilities of success and magnitudes of good to be achieved if an effort is successful prevents certain options, such as becoming an anti-capitalist revolutionary, from being on the table in the first place, since there seems to be no way of estimating the value of bringing about an unrecognizably different sort of society, or the probability of succeeding in doing so.

In fact, however, neither Herzog's nor Srinivasan's arguments provide us with reasons to think that effective altruists cannot in principle endorse individuals devoting resources and/or time to efforts to promote institutional reform that there seem to be good moral reasons to support. It might turn out, for example, that the best available evidence suggests that at least some people can maximize the expected value (where reducing suffering and increasing happiness are, consistent with EA2, given significant, though not necessarily exclusive, weight in estimating the value of outcomes) of their career choices by seeking political office and attempting to enact important policy changes.<sup>39</sup> Or, it might be that the expected value of one's devoting time to attempting to persuade others to join an activist cause

<sup>38</sup> Herzog, '(One of) Effective Altruism's Blind Spot(s)'.

<sup>39</sup> As Srinivasan herself notes ('Stop the Robot Apocalypse'), MacAskill discusses this possibility at some length (*Doing Good Better*, pp. 89–95).

promoting important institutional reforms is high enough that it is among the most effective ways that one might improve the world. In principle, we could produce at least a rough estimate of the expected value of an effort like this by testing how effective people with, for example, different educational backgrounds, skill profiles, etc., are in persuading others to join relevant activist causes, and estimating what is achieved, on average, by a relevant number of additional people joining such causes. This is, in fact, precisely the sort of investigation that effective altruists tend to think ought to be done in order to test potential ways of improving the world for effectiveness, so that we can all make more informed choices that will, hopefully, have a greater positive impact on the lives of the global poor.

But the sort of investigation that I described clearly does not simply take existing institutional structures for granted; nor does it treat individuals as isolated atoms who cannot attempt to work together to bring about social change. One of the ways that we can affect the world for the better is by engaging with others who might be persuaded to join us in efforts to do so. But, for effective altruists, whether this is a well-justified use of our time and resources depends, at least in part, on how successful it is likely to be. Working with others who are committed to a cause, organizing new collective efforts where they do not already exist, and engaging in efforts to persuade others to join in collective efforts to promote justice via institutional reform are options to be considered. But, as individuals, we often cannot be certain that enough others will be willing to join in any particular collective effort for that effort to be likely enough to succeed to justify investing substantial time and resources in it. And, typically, we cannot simply force others to contribute to the collective efforts that we favour when we find that they are unwilling to contribute voluntarily. Even in cases in which we could do this, in the sense that we would succeed if we tried, it would, perhaps with some unusual exceptions, be wrong to do so. Given these practical and moral constraints, it is always, for effective altruists, an open question what the best ways available to improve the world might be for any particular individual, and this question is always to be settled, to the extent that it can be, by examining the best evidence available, and producing new evidence where it is needed.

It might be objected that my claim that the likelihood that enough others will contribute to a collective effort to promote institutional reform is a factor that should be given significant weight by individuals when determining whether to contribute themselves frames the question in an objectionably individualistic way.<sup>40</sup> Perhaps individuals

<sup>40</sup> I am grateful to an anonymous reviewer for suggesting that I discuss this objection.

should, instead, structure their deliberations around fundamentally collective reasons or obligations. If one is a member of a group that is collectively obligated to reform institutions that are among the root causes of poverty,<sup>41</sup> then, it might be argued, she has a very strong reason (perhaps an obligation) to do her part in promoting the required reforms by devoting time and resources to the relevant efforts, regardless of whether the other members of the group will do their parts as well. This reason, it might be argued, will at least typically be stronger than any welfare-based reasons she has to direct the same time and resources to, for example, organizations such as the Against Malaria Foundation or GiveDirectly. If this view is correct, then the effective altruist claim that whether an individual has most reason to direct her time and resources to institutional reform efforts will typically depend, to a significant extent, on what others are likely to do, is mistaken.

It is implausible, however, that one will generally have stronger reasons to direct time and resources to institutional reform efforts than to direct the same time and resources to organizations that reliably benefit the global poor, even in cases in which there are strong reasons to believe that, no matter what one does oneself, relatively few members of the collectively obligated group will contribute to the relevant institutional reform efforts. This is generally acknowledged even by those who believe that collectives can have obligations, and that these collective obligations bear on what individual members of the obligated collectives ought to do. Consider, for example, the following case:

*Drowning Children:* Five children are in danger of drowning. All five can be saved if, and only if, A, B, and C work together to paddle a nearby canoe to the children. The group consisting of A, B, and C, then, is collectively obligated to rescue the children. If fewer than three people attempt to rescue the children with the canoe, however, they will arrive too late, and all 5 children will drown. A has very good reason to believe that C will not do his part to contribute to the satisfaction of the collective obligation, regardless of what A and B do. B, on the other hand, is committed to doing her part, regardless of what the others do. A has to decide whether to join B in the canoe.

<sup>41</sup> I assume that collectives can have obligations in their own right. This question has generated much recent discussion; see, for example, Bill Wringe, 'Collective Obligations: Their Existence, Their Explanatory Power, and their Supervenience on the Obligations of Individuals', *European Journal of Philosophy* 21 (2014), pp. 472–97, and Stephanie Collins and Holly Lawford-Smith, 'Collectives' and Individuals' Obligations: A Parity Argument', *Canadian Journal of Philosophy* 46 (2016), pp. 38–58.

If she does not, then she can run down the shore and reach two life preservers in time to throw them and save two of the children.

It seems clear that A ought to run to the life preservers, despite the fact that this entails failing to do her part to contribute to the satisfaction of the collective obligation.<sup>42</sup> Because she has very good reason to believe that the collective obligation will not be satisfied even if she does her part, she has reason to do something else that can be expected to help at least some, though not all, of those to whom the collective is obligated.

Proponents of the institutional critique might object that joining efforts to promote institutional reform that would address the root causes of poverty is, in a morally important respect, quite unlike joining B's almost certainly futile canoe rescue effort. Specifically, success in institutional reform efforts is a matter of degree,<sup>43</sup> whereas in *Drowning Children* it is stipulated that the only possible outcomes of the canoe rescue effort are rescuing all of the children and rescuing none of them. It can be granted that in cases like *Drowning Children*, A ought to run to the life preservers rather than joining B in the canoe, while insisting that in cases in which the success of an obligatory collective effort is a matter of degree, and in which the expected degree of success will, at least roughly, vary with the number of contributors, individuals will at least typically have stronger reasons to contribute to the collective effort than to benefit people in need in other ways, regardless of what the other members of the obligated collective can be expected to do.

This view, however, also generates quite counterintuitive results. Consider the following case:

*Collective Pain Infliction*: D is hooked up to a machine with 100 switches. The switches fluctuate between being on and off, depending on the actions of ten people, including E, F, and G. At any given time, D is in an amount of pain, caused by the machine, that corresponds to the number of switches that are on; the more switches are on, the more pain he is in.<sup>44</sup> The other seven of the ten people who can

<sup>42</sup> It might be objected that in appealing to this kind of case, I am doing one of the things that some proponents of the institutional critique argue against, namely treating our obligations to the global poor as analogous to our obligations to drowning children. But my argument here does not rely on the assumption that the cases are analogous in all morally relevant respects. The case functions only as a counterexample to the claim that individuals generally have stronger reasons to do their part towards the satisfaction of a collective obligation than to benefit people in need in other ways, regardless of what other members of the obligated collective can be expected to do.

<sup>43</sup> For example, a collective effort to enact a fully just policy might lead to the enactment of a policy that is an improvement on the status quo, though not fully just.

<sup>44</sup> Assume that for each switch that is turned on, D suffers a perceptible increase in pain.

affect how many switches are on often turn a fair number of switches on, despite the fact that it is clearly wrong for them to do so, since it allows them to acquire certain benefits. E, F, and G never turn any switches on, and believe that the group consisting of all ten of them is collectively obligated to ensure that no switches are ever on. The three, however, have very good reason to believe that any effort that they might make, either by force or persuasion, to get the others to keep all of the switches off, will fail. F and G have, however, pooled some of their resources together in order to buy a replacement machine that causes slightly less pain for each switch that is turned on. E has \$10,000 that she can either add to F and G's pool, in which case they can buy a slightly better replacement machine that causes even less pain per switch, or use to buy a lifetime supply of pain medication to give to D. The medication would have no side effects, and would reduce D's pain by more than the better replacement machine would.

Putting the money towards the better replacement machine would improve things at the level of the root cause of D's suffering in a way that providing the medication would not. Nonetheless, it seems to me fairly clear that E has stronger moral reason to provide the medication than to fund the purchase of the better replacement machine. If this is correct, then an increased emphasis on framing issues such as global poverty as generating fundamentally collective obligations will not provide support for the view that individuals will generally have stronger reasons to contribute to institutional reform efforts than to contribute to other efforts that can be expected to benefit the global poor.

Perhaps, however, proponents of the institutional critique would argue that it is not the failure to think about global poverty as generating fundamentally collective obligations that explains why effective altruism is objectionably biased against efforts to promote institutional reform. Instead, they might claim that there is something problematic about the emphasis that effective altruists place on estimating the effects of potential courses of action on the welfare or quality of life of the people who are affected by the relevant efforts. They might argue, for example, that effective altruists tend to take improvements in welfare or quality of life for the global poor to be more important, morally speaking, than they in fact are, and that, relatedly, they ignore, or at least underestimate the significance of, other values that provide strong reasons to favour efforts to promote institutional reform over efforts to increase welfare or quality of life by other means, even if the most promising available means of attempting to promote institutional reform seem quite unlikely to succeed, and so



quite unlikely to make the lives of the global poor better in welfare or quality of life terms.

It is true that effective altruists will typically not recommend that individuals devote much of their time and resources to efforts to promote institutional reform that appear to have little or no chance of succeeding, and therefore appear to have little or no chance of actually improving the lives of the global poor in welfare or quality of life terms. This is simply an effect of effective altruists' commitment to EA2 and EA3. Intuitively, however, this seems not only unobjectionable, but clearly appropriate. Even if the success of an anti-capitalist revolution would be a good thing, if there are not nearly enough committed revolutionaries to make the prospects for a successful revolution greater than infinitesimal, and there are good reasons to expect this to continue to be the case for the foreseeable future, then it seems clear that joining the revolutionary cause would not be the morally best use of one's time, energy, and resources. For effective altruists, whether we are, as individuals, in circumstances with that structure regarding any particular effort to promote institutional reform whose success would be a good thing is an empirical question, and ought to be settled, to the extent that it can be, by examining the best available evidence, as well as seeking to develop new evidence.<sup>45</sup>

Proponents of the institutional critique might be interpreted as rejecting the claim that we should rely on and gather empirical evidence about what different kinds of efforts tend to accomplish, and how likely they are to succeed. That is, they might be understood as rejecting EA4 in favour of the view that we ought to aim at institutional changes regardless of what any evidence suggests about, for example, the likelihood that such efforts will succeed. Perhaps they believe that it is always morally better to attempt to address the root causes of large-scale moral problems such as global poverty than to contribute

<sup>45</sup> I have discussed anti-capitalist revolution as an example of institutional change, primarily because several proponents of the institutional critique discuss it, and individual efforts in support of it, favourably. But of course there are far less radical institutional changes that proponents of the institutional critique might advocate supporting, and to the extent that these reforms are more likely to be achievable, would involve fewer transition costs, come with fewer risks of unintended and unforeseen negative effects, and have other features that, at least in themselves, make the expected value of pursuing them greater, effective altruists will be more likely to support individuals dedicating time and resources to efforts to bring them about. Nothing in my argument, and nothing in the core commitments of effective altruism as I have presented them, rules out the possibility that effective altruists should all in fact be dedicating most of their time and resources to efforts to promote institutional reform. My claim is merely that effective altruists are correct that whether this is what individuals have most reason to do depends on what the available evidence suggests will best promote the values that ought to guide our response to global poverty and other morally important issues.

to alleviating the symptoms, even if there are strong reasons to believe that one's efforts to address the root causes will be unsuccessful.

This, however, is a rather implausible view. Surely we should consider available evidence about what different things that we might do are likely to accomplish when deciding how to contribute to addressing global poverty. Holding that individuals should focus their efforts on addressing root causes, and therefore on institutional change, regardless of what such evidence suggests, is, as Jeff McMahan puts it, 'rather like condemning a doctor who treats the victims of a war for failing to devote his efforts instead to eliminating the root causes of war'.<sup>46</sup>

A better response on behalf of proponents of the institutional critique is to argue that effective altruists operate with an overly narrow conception of what counts as good evidence for the (potential) effectiveness of an effort to benefit the global poor.<sup>47</sup> For example, it is sometimes suggested that effective altruists' emphasis on relying on evidence about effectiveness and probability of success leads to a bias in favour of efforts the effects of which are easily measurable and quantifiable. To some extent, effective altruists invite this charge, since a significant proportion of their public outreach efforts highlight, often exclusively, their support for organizations whose programmes have been tested for effectiveness in randomized control trials (RCTs), and performed well. This is, I suspect, because the primary aim of such public outreach efforts is to persuade individuals to accept both that there are strong reasons to donate to organizations that benefit people more effectively, rather than to others that are less effective, and that there is often evidence available about the relative effectiveness of different organizations. Many people believe that there is no way to know how much good different charitable organizations achieve with the donations that they receive, and this often leads them to believe either that there is nothing objectionable about donating to whatever organizations they might, for whatever reason, prefer, or worse, that there are not strong reasons to donate at all. Highlighting that RCTs can be used to estimate, in a reasonably reliable way, what is achieved by particular organizations' efforts is an effective way of countering these widely held views. Effective altruists' efforts here have, however, contributed to the impression that they believe that nearly all of our efforts to, for example, combat global poverty should run through organizations whose programmes have been shown by RCTs to produce significant results.

<sup>46</sup> Jeff McMahan, 'Philosophical Critiques of Effective Altruism', *The Philosophers' Magazine* 73 (2016), pp. 92–99, at 97.

<sup>47</sup> I am grateful to an anonymous reviewer for suggesting that I discuss this response.

The concern that effective altruists have overemphasized the importance of relying on quantifiable evidence of the kind that RCTs can provide seems to me legitimate. There are, in my view, important questions about the extent to which we ought, or ought not, to prioritize supporting efforts that have been shown, via RCTs or similarly rigorous methods, to have measurable beneficial results, over efforts that cannot be tested for effectiveness in these ways. To the extent that some effective altruists are quick to be dismissive of efforts that cannot be tested using RCTs or similar methods, they can, I think, be criticized for having an overly narrow conception of how we should go about deciding where to direct our time and resources. But I think that it would be a mischaracterization to suggest that the movement as a whole is especially vulnerable to this criticism.<sup>48</sup> As even some proponents of the institutional critique point out, both Singer and MacAskill discuss certain institutional reform efforts quite favourably, despite the fact that their expected effects and likelihood of success cannot be estimated using RCTs or other similar methods. Effective altruist organizations like GiveWell are involved in efforts, via the Open Philanthropy Project, to think about what might be done in a range of policy areas, as well as what we might have reason to do in response to various ‘global catastrophic risks’, such as anthropogenic climate change and threats from artificial intelligence.<sup>49</sup> And the Centre for Effective Altruism and the Future of Humanity Institute (which share several team members, including MacAskill and Toby Ord) partner, via the Global Priorities Project, in thinking about approaches to existential risk and cause prioritization among morally important efforts.<sup>50</sup> Neither (at least most of) the policy initiatives that the Open Philanthropy Project considers as possible candidates for effective altruist support, nor, certainly, potential approaches to mitigating catastrophic and existential risks, can, however, be tested for effectiveness using RCTs or similar methods. It seems clear, then, that effective altruists are, generally speaking, very much open to thinking that we might have good reasons to support efforts that have not been shown to be effective using RCTs or similar methods.<sup>51</sup> It is very difficult to think in a clear and systematic way

<sup>48</sup> Though, again, much public outreach on behalf of effective altruism has a tendency to suggest otherwise.

<sup>49</sup> See <<http://www.openphilanthropy.org/research/cause-reports#Policy>>.

<sup>50</sup> See <<http://globalprioritiesproject.org/about-us/>>.

<sup>51</sup> The case of effective altruist support for efforts to think about how we might effectively mitigate existential risk is particularly interesting to consider in relation to the institutional critique. Their support for such efforts shows that effective altruists clearly think that there can be reasons to pursue efforts that have a low probability of doing good, but would do a very large amount of good if they turn out to be successful (and, in the case of mitigating existential risks, necessary), even in preference to efforts

about how we ought to prioritize, in allocating time and resources, among a wide range of efforts, when in many cases the probabilities of success, benefits of success, and other relevant factors are impossible to quantify and can be estimated only in the roughest possible way. But no one, effective altruist or otherwise, can avoid attempting to confront this challenge as best they can. It seems to me that effective altruists have at least attempted to engage with these challenging issues, though of course to what extent they have succeeded in arriving at well-justified conclusions remains, I think, largely an open question.

It is less clear to me, however, that proponents of the institutional critique recognize just how difficult these issues are. Their tendency to suggest that it is clearly a mistake to take seriously the possibility that directing money to organizations such as the Against Malaria Foundation is among the things that we have most reason to do in response to global poverty, and to advocate institutional reform efforts as the clearly preferable alternative, all without saying anything about potential differences in probabilities of success, or how we ought to think about the moral importance of these potential differences, can appear to indicate that they think that these issues are simply not relevant to how we ought to decide where to direct our efforts. Perhaps proponents of the institutional critique can be interpreted as holding that it is simply an empirical fact that promoting institutional reform is always, or at least nearly always, the most effective means by which individuals can improve the lives of the global poor in relevant respects, including increasing welfare or quality of life. This view, however, seems implausible, and although some critics point to cases in which efforts to improve the lives of some of the global poor in ways other than via institutional reform have had bad consequences, including degrading important government-run institutions,<sup>52</sup> it is effective altruists themselves, and not at least many of their opponents, who are committed to carefully taking into account all relevant empirical evidence, including the evidence generated by cases in which efforts to

that are virtually certain to do a fair bit of good. At least certain efforts to bring about institutional reform seem to share these features. This suggests that effective altruists should be open to the possibility that these efforts, too, should be prioritized over higher probability, lower payoff efforts. I have suggested that, despite the fact that their public outreach efforts can have a tendency to suggest otherwise, effective altruists are, at least for the most part, consistent on this point. I suspect that many proponents of the institutional critique, on the other hand, will find effective altruist support for efforts to mitigate existential risk to be quite misguided. It's unclear, however, what principled basis there might be for holding that the institutional reform efforts that they favour should necessarily be prioritized both over efforts like those of organizations like the Against Malaria Foundation or GiveDirectly, and over efforts to mitigate existential risks.

<sup>52</sup> See, for example, Clough, 'Effective Altruism's Political Blind Spot'.

improve the lives of the global poor that may have seemed promising turned out to have unanticipated bad effects, when deciding how to allocate their time and resources.<sup>53</sup> If it turned out that the best available evidence suggested that devoting time and resources to efforts to promote institutional reform is the most effective way to improve the lives of the global poor in relevant ways, there is nothing in the core commitments of effective altruism that would preclude effective altruists from encouraging individuals to direct their time and resources to such efforts. If the institutional critique is primarily premised on the empirical claim that promoting institutional reform is typically the best means available to individuals of improving the lives of the global poor in relevant respects, then it is not in principle inconsistent with effective altruism, but rather simply asserts that effective altruists tend to make mistakes when applying their core commitments to the question of what individuals have the most reason to do given the facts on the ground.

Lichtenberg, Herzog, Mills and Srinivasan all seem to understand their critiques as providing reasons to reject effective altruism itself, rather than merely reasons to doubt their applications of their core commitments to the practical questions that individuals face in the real world. They seem to hold, then, that there would be reasons to reject the recommendations of effective altruists regarding where individuals ought to direct their time and resources, even if it were true that effective altruists were correctly identifying the ways of employing one's resources that would do the most to improve the welfare or quality of life of the global poor. In order to defend this view, proponents of the institutional critique would need to successfully defend at least one of the following two closely related claims. The first is that there are important values that provide us with strong reasons to prefer engaging in efforts to promote institutional reform, even if the available evidence suggests that such efforts will do less, and perhaps significantly less, to improve the welfare or quality of life of the global poor than other things that one might do instead (such as contributing to the Against Malaria Foundation or GiveDirectly). The second is that effective altruists significantly overestimate the moral importance of the welfare or quality of life improvements that the efforts of, for example, the Against Malaria Foundation or GiveDirectly achieve.

What values might proponents of the institutional critique point to in order to defend the first claim, and perhaps lend support to the second? In other words, how might they explain why we should prefer directing

<sup>53</sup> See, for example, MacAskill's discussion of the PlayPump (*Doing Good Better*, pp. 1–10).

our time and resources towards efforts to promote institutional reform even when the best available evidence suggests that doing so is likely to do less, and perhaps significantly less, to improve the welfare or quality of life of the global poor than would directing the same time and resources towards efforts recommended by effective altruists? What values could be important enough that they ought to be given priority over improving the welfare or quality of life of the global poor, when what is at issue is how we ought to respond to the morally troubling facts about global poverty? When the questions are put in these terms, it is, it seems to me, difficult to imagine what a compelling answer might look like. But these are the questions that proponents of the institutional critique must provide an answer to in order to make the case that effective altruism's core commitments are in some way objectionable.

In somewhat different contexts, it is often argued that properly respecting persons can require acting in ways that do not maximally promote welfare. For example, properly respecting a person can require avoiding paternalistically interfering with her autonomous choices, or refraining from harming her in order to bring about greater aggregate benefits for others. In these cases, it seems at least plausible that the value of respect provides reasons that override the reasons that we have to increase welfare. Might respect for the global poor call for directing our time and resources towards efforts to promote institutional reforms required by justice, even if doing so will do less to increase welfare or quality of life for those very people? Perhaps the proponents of the institutional critique would argue that respect for the global poor requires acting in a way that seeks to provide them with everything that they are entitled to as a matter of justice, rather than merely aiming to alleviate some of the suffering caused by their victimization by an unjust global institutional structure. If this were correct, then (assuming that an anti-capitalist revolution would be necessary in order to achieve complete justice for the global poor) becoming an anti-capitalist revolutionary may be required as a matter of respect for the global poor, despite the fact that doing this might do less to improve their lives than, for example, working for an investment bank and donating the majority of one's salary to the Against Malaria Foundation.

It is unclear, however, why exactly we might think that respect for the global poor could require that we devote time and resources to efforts that there are strong reasons to think cannot succeed. In other cases in which respect seems to call for acting in a way that does not maximally promote welfare, it is clear that those to whom we are required to show respect would have at least some grounds

for complaint against the problematically disrespectful conduct. If I attempt to interfere paternalistically with a person's choices, for example, she can complain that I am seeking to undermine her autonomy, and that such attempts are objectionably disrespectful.<sup>54</sup> But it is unclear what kind of complaint individuals who are the victims of global injustice might have against an individual who *would support efforts to create just global institutions by devoting time and resources, if such efforts were sufficiently likely to succeed*, but chooses instead to attempt to increase the welfare of the victims of global injustice by other means that are likely to be more effective, since there are simply not enough people working for just global institutional reform. Indeed, it seems that the global poor would have good reason to prefer that well off beneficiaries of global injustice *avoid* devoting time and resources to efforts to promote institutional reform (even reform required by justice) that are extremely unlikely to succeed, and so extremely unlikely to improve their lives. It does not even seem implausible to think that devoting resources to institutional reform efforts, when those resources could do much more to improve the lives of the global poor if directed elsewhere, would itself be objectionably disrespectful of the victims of global injustice.<sup>55</sup>

But perhaps the proponents of the institutional critique would argue that the effective altruists' primary mistake is not neglecting other values, but instead overestimating the moral importance of the increases in welfare or quality of life that can be achieved by, for

<sup>54</sup> Concerns about paternalism may give us some reason to prefer directing resources to, for example, GiveDirectly, as opposed to other organizations that efficiently promote the welfare of the global poor (see Emma Saunders-Hastings's contribution to the *Boston Review* forum on 'The Logic of Effective Altruism': <<https://bostonreview.net/forum/logic-effective-altruism/emma-saunders-hastings-response-effective-altruism>>).

<sup>55</sup> A further issue that proponents of the institutional critique have not discussed is that even if they are successful, the efforts to promote institutional reform that they advocate would not improve the lives of many of the people whose lives would be improved by directing resources towards organizations recommended by effective altruists. Achieving institutional reforms that make the world more just takes a great deal of time, in particular at the global level, and this is especially so in cases in which the reforms sought will be vigorously opposed by many who benefit from the injustices entrenched within the status quo. Proponents of the institutional critique surely recognize this. But an important consequence of this fact is that advocates of the institutional critique are not merely advocating that we attempt to improve the lives of some single group of people by pursuing a lower probability, higher reward option, rather than by pursuing an alternative, high probability, lower reward option. They are instead advocating that we refrain from taking high probability steps to alleviate the suffering of today's global poor, in order to pursue low probability, potentially high reward efforts to improve global institutions, so that *different people*, sometime in the future, are able to live under more just institutions than might otherwise exist. Whether or not we might be justified in prioritizing efforts to make global institutions more just for future people over improving welfare or quality of life for the current global poor, it seems clear that we cannot be required to do so *as a matter of respect for the current global poor*.



example, the charitable organizations that they tend to recommend. The thought here would be that we have strong reasons to focus our attention and resources on efforts to promote institutional change, even if the prospects of success are very low, because success would advance values that we have strong reasons to care about, while success in increasing welfare or quality of life that is not achieved via institutional change is of much less moral importance than effective altruists typically take it to be. This view might be suggested by Herzog's claim that '[w]hat matters for a good human life, in which basic needs can be justified and individuals have some degree of autonomy, is that . . . institutions and practices function to the advantage of everyone'.<sup>56</sup> If we interpret this claim as implying that having institutions that meet certain conditions is a *necessary condition* for people who are currently unjustly disadvantaged by the global order coming to have good (or at least better) lives in morally relevant respects, then its truth would suggest that there may be less reason to attempt to increase welfare or quality of life in ways other than by promoting institutional change than effective altruists tend to think.

There is a view that has a structure that is at least somewhat similar that is familiar in political philosophy. This is the broadly Rawlsian view that justice is at least primarily a matter of what the institutions of the 'basic structure of society'<sup>57</sup> are like, and less, or at least less directly, a matter of how individuals act and are motivated within those institutions. On this view, individuals' efforts to improve the lives of those who are victims of unjust institutions, such as voluntarily directing resources to such people, either do not in fact do much, if anything, to eliminate injustice, or, more strongly, cannot, as a conceptual matter, make an unjust society any less unjust.<sup>58</sup> From the perspective of justice, then, Rawlsians typically hold that there are not strong reasons (or any reasons) for individuals to directly aid the badly off, despite the fact that there are strong reasons to contribute to efforts to reform the institutions of the basic structure when they are unjust, so that those institutions will in turn improve the lives of the badly off.

It seems to me that this view is difficult to defend, since it requires holding that even very substantial voluntary redistribution of resources from the unjustly advantaged to the unjustly disadvantaged would not make things *any better* in respect of justice, even if it had the

<sup>56</sup> '(One of) Effective Altruism's Blind Spot(s)'.

<sup>57</sup> See, for example, John Rawls, *A Theory of Justice*, rev. edn (Cambridge, MA, 1999), pp. 6–10.

<sup>58</sup> For the latter interpretation, see G. A. Cohen, *Rescuing Justice and Equality* (Cambridge, MA, 2008), pp. 126–7.

effect of dramatically improving the quality of life for the (previously) badly off.<sup>59</sup> And this requires holding that a concern about welfare or quality of life provides no part of the appropriate ultimate grounds of our concern about justice. This implication, however, might seem at least somewhat more palatable once we recognize that the view concerns the content of our justice-based reasons only, and not the content of all of the moral reasons that we might have to directly aid the badly off. It is open to those who hold that we have justice-based reasons to attempt to reform unjust institutions, but no such reasons to directly aid the victims of unjust institutions, then, to hold that we nonetheless have strong (non-justice-based) moral reasons to directly aid the badly off, and that these reasons might make it the case that, all things considered, this is where we ought to direct our resources in cases in which efforts to reform unjust institutions are sufficiently unlikely to succeed.

The critics of effective altruism, however, seem to think not merely that we have no *justice-based* reasons to provide aid to the global poor in some of the ways that effective altruists tend to endorse, but also that whatever moral reasons we have to provide such aid are outweighed by the reasons to focus our efforts on promoting institutional reform. But in order for this to be the case, it would seem to have to be the case not only that prospects for improved welfare and quality of life among the global poor do not provide us directly with justice-based reasons to act, but also that such prospects do not provide us with strong enough moral reasons of any kind to outweigh the reasons provided by whatever other considerations support promoting institutional reform, even in cases in which such efforts have relatively little prospect of improving welfare or quality of life. On such a view, however, welfare or quality of life improvements would seem to have to count for relatively little, morally speaking. But this, it seems to me, is a clearly unacceptable implication of a view about the content of the reasons that we have to act in response to the facts on the ground concerning global poverty. Accepting it would, it appears, require rejecting EA2; but if the institutional critique depends on the rejection of EA2, then we have good reason to reject the critique, since EA2 is a fairly weak claim that seems clearly correct.<sup>60</sup>

<sup>59</sup> See Brian Berkey, 'Against Rawlsian Institutionalism about Justice', *Social Theory and Practice* 42 (2016), pp. 706–32; 'Double Counting, Moral Rigorism, and Cohen's Critique of Rawls: A Response to Alan Thomas', *Mind* 124 (2015), pp. 849–74; 'Obligations of Productive Justice: Individual or Institutional?', *Critical Review of International Social and Political Philosophy* (forthcoming).

<sup>60</sup> An alternative for the proponent of the institutional critique would be to hold that welfare or quality of life improvements do provide strong reasons, but that whatever other kinds of reasons support promoting institutional reform are *extremely weighty*, even in cases in which the prospects for success are minimal, so that they continue to outweigh

#### IV. THE (ILLUSORY) APPEAL OF THE INSTITUTIONAL CRITIQUE

What might explain why proponents of the institutional critique seem willing to accept that there are not particularly strong reasons to direct resources to organizations such as the Against Malaria Foundation and others often recommended by effective altruists? I suspect that part of the answer is that some are prone to what Peter Unger has called ‘futility thinking’ about the significance of a single individual’s contributions to such organizations.<sup>61</sup> For those in the grip of futility thinking, sacrifices that would help a small number of people in a much larger group are viewed as accomplishing nothing more significant than ‘removing a mere drop of trouble from a whole sea of suffering’,<sup>62</sup> even when the effects of such sacrifices include the saving of a number of lives. Since the number of people suffering from global poverty and its effects is in the billions, it is easy to think that the only thing that could improve things in a way that we would have strong reasons to care about would be substantial institutional reform, and easy to think that a single person’s contributions to organizations like the Against Malaria foundation, which, for a typical well off person (as opposed to, say, a billionaire), may, over a lifetime, save a few dozen people’s lives, do not matter much at all.<sup>63</sup> And if one thinks these things, then it

the strong reasons provided by clearly achievable welfare or quality of life improvements. This view is not, strictly speaking, inconsistent with EA2. It does not, however, seem any more plausible, since it requires denying that the strength of the reasons that we have to contribute to an effort diminish as the prospects for success decrease.

<sup>61</sup> Peter Unger, *Living High and Letting Die: Our Illusion of Innocence* (New York, 1996), p. 75.

<sup>62</sup> Unger, *Living High and Letting Die*, pp. 75–6.

<sup>63</sup> See, for example, Robert Simpson’s claim that ‘even where our giving can alleviate some people’s hardships in the short term, there’s little reason to be confident – without systematic change – that our giving will have any real effect on suffering in the grander scheme of things’ (‘Moral Renegades’), and Tracy Isaacs’s claim that ‘[i]n the face of pressing issues of enormous scale and scope, such as child poverty, animal suffering, and climate change, individual effort will rarely make a marked difference except for the likes of Ted Turner and others who have a billion dollars to give’ (Tracy Isaacs, ‘The Most Good We Can Do: Comments on Peter Singer’s *The Most Good You Can Do*’, *Journal of Global Ethics* 12 (2016), pp. 154–60, at 155–6). Notice that Simpson’s explicit suggestion is that, for some reason, alleviating some people’s hardships (by, for example, donating enough money to cure ten people of trachoma-caused blindness) might not count as a *real* effect on suffering in the grander scheme of things. But, if taken literally, this cannot be what he intends to say, since surely the reductions in suffering that we can bring about by donating to organizations recommended by effective altruists are no less real than similar reductions that could be brought about via institutional change. What Simpson must mean here is that successful institutional change would reduce suffering *much more* than a single individual’s donations to effective altruist-recommended organizations could. And this is clearly true, though it is not clear why we might think that it is relevant to what any particular individual should do with her resources. After all, no individual can bring about institutional change on her own.

is easy to see why one would believe that we do not have especially strong reasons to give to the Against Malaria Foundation, and should direct whatever efforts to improve the world that we do make towards institutional reform.

As Unger effectively argues, however, futility thinking is deeply misguided.<sup>64</sup> If I were in a position to save the lives of a few dozen people in serious danger right before my eyes, and it would cost me, say, 20 per cent of my lifetime earnings to do so, I would have extremely strong, and even, most would agree, decisive moral reason to make the sacrifice necessary to save those lives. In addition, I would have no less reason to make this sacrifice if there were thousands, or even millions, more people in danger whose lives I could not save. The moral importance of saving the lives of those whom I am in a position to save does not diminish simply because many others, whom I cannot save, happen also to be in similar danger. There is, moreover, no reason to deny that this is the case when those whom I could save happen to be distant strangers in danger of being stricken with malaria rather than people who happen to be nearby. It is simply a mistake to think that I do not have strong reasons to save lives by giving to the Against Malaria Foundation, merely because there are millions of others who will not be saved regardless of what I do. If I can save several dozen lives by giving, or instead direct my efforts at institutional reform, then if the latter is likely to have little or no effect due to the non-cooperation of others, I will have most reason to do the former.<sup>65</sup>

A second, and perhaps more important, reason why many may find the institutional critique appealing is that it can appear to offer us the prospect of insisting that a great deal ought to be done to alleviate the plight of the global poor, while de-emphasizing (at least when compared to effective altruism) the issue of how much of their resources well off people ought to be sacrificing. It is appealing to think that the real work of addressing global poverty and related problems must, as a matter of principle, be done by institutions, rather than by individuals, so that our responsibilities as individuals are limited to doing our part to promote the development of the necessary institutions. If this view is correct, we might think, then we do not need to worry too much that we might be acting wrongly when we spend significant amounts of money pursuing projects and interests that we care about, at least so long as we engage in enough political activity in support of the necessary institutional change (e.g. voting for the right candidates, attending

<sup>64</sup> Unger, *Living High and Letting Die*, pp. 80–2.

<sup>65</sup> For a similar point, see McMahan, 'Philosophical Critiques of Effective Altruism', p. 97.

rallies, organizing, and perhaps even contributing some money to relevant political efforts).

This explanation of the appeal of the institutional critique seems to be supported by the fact that a number of those who endorse that critique also endorse what would at least seem, on the surface, to be a very different line of criticism of effective altruism. This is that the commitments of effective altruism require that we prioritize, at least to a fairly significant extent, acting in the ways that will improve the world the most over, for example, pursuing personal projects that would enrich our own lives, advancing the interests of those about whom we care deeply, or helping those among the needy to whom we have special connections.<sup>66</sup> A commitment to effective altruism, according to this criticism, would leave individuals alienated from the features of their lives that make life worth living, and so it is objectionably demanding to require them to develop such a commitment.<sup>67</sup>

If one thinks that requiring individuals to set aside their personal commitments, at least to some significant extent, in order to contribute to the efforts to alleviate global poverty that seem likely to do the most good is objectionably demanding, and yet believes that, in some important sense, it is a moral imperative to address global poverty and the suffering that goes with it in a serious way, it is tempting to think that the primary locus of responsibility must be institutional, rather than individual. But in fact attempting to direct responsibility to institutions can do little, if anything, to reduce the moral burdens on individuals in conditions in which well-functioning institutions do not yet exist.<sup>68</sup> However demanding it is for an individual who cares deeply about the members of her (well off) family, charities to which she feels a special connection (such as, for example, one that trains guide dogs for the blind),<sup>69</sup> and personal projects of the sort that many well

<sup>66</sup> See, for example, both Lichtenberg's ('Peter Singer's Extremely Altruistic Heirs') and Srinivasan's ('Stop the Robot Apocalypse') criticisms of MacAskill's discussion of his decision not to donate to the Fistula Foundation, despite the fact that he had met women suffering from fistulas.

<sup>67</sup> This criticism of effective altruism derives from Bernard Williams's well-known critique of utilitarianism; see 'A Critique of Utilitarianism', *Utilitarianism: For and Against*, ed. J. J. C. Smart and Bernard Williams (Cambridge, 1973), pp. 75–150. For articulations of this criticism directed against effective altruism, see John Gray, 'How and How Not to Be Good', *New York Review of Books* (May 2015): <<http://www.nybooks.com/articles/2015/05/21/how-and-how-not-to-be-good/>>; Nakul Krishna, 'Add Your Own Egg', *The Point Magazine* (2016): <<http://thepointmag.com/2016/examined-life/add-your-own-egg>>.

<sup>68</sup> Of the proponents of the institutional critique, only Herzog explicitly recognizes this ('Can Effective Altruism Really Change the World?'; '(One of) Effective Altruism's Blind Spot(s)').

<sup>69</sup> Srinivasan appears to endorse the view that it's permissible for a person to give to a seeing-eye dog charity if one feels especially connected to its cause ('Stop the Robot Apocalypse').

off people pursue, to deprioritize those values in favour of directing resources to the highly effective charities recommended by effective altruists, it will be equally demanding for them to deprioritize those values in favour of directing resources to anti-capitalist revolutionary efforts, or any other political efforts that can succeed only with massive collective investments of time and resources.

Proponents of the institutional critique typically seek to present themselves as more thoroughgoing, and more principled, champions of justice for the global poor than are their effective altruist opponents. Achieving meaningful and positive institutional change is difficult, resource intensive, and requires substantial participation and cooperation among those committed to bringing it about. Proponents of the institutional critique know this, and insist, in no uncertain terms, that we must strive to achieve it nonetheless. Given this insistence, we might expect them to advocate institutional change-focused versions of some of effective altruism's main initiatives. For example, they could suggest that well off individuals ought to commit themselves to devoting a significant percentage of their income to promising efforts to bring about important institutional reforms that would benefit badly off people.<sup>70</sup> Or they could suggest establishing an institutional change-focused organization to gather evidence about which efforts to promote institutional change individuals should contribute to, taking into account, for example, what the efforts would be likely to achieve if successful, how likely they are to succeed, and whatever other factors they might think relevant to determining which efforts there are the strongest reasons to support.

Recommendations of this sort are, however, conspicuously absent from all of the pieces in which the institutional critique features that I have discussed.<sup>71</sup> And this makes it difficult not to suspect that proponents of that critique are, to a significant extent, motivated by the desire to avoid accepting that well off individuals ought to be making significant sacrifices in order to contribute to addressing global suffering and injustice, rather than by a commitment to global justice that extends beyond that of effective altruists. Indeed, since these motives are, at least to a significant extent, in direct conflict, the more

<sup>70</sup> The effective altruist organization Giving What We Can (<<https://www.givingwhatwecan.org/>>) asks people to pledge to give 10 per cent of their income each year to organizations that they recommend as especially effective providers of aid to the global poor.

<sup>71</sup> Jennifer Rubenstein has a helpful discussion of some of the questions that those sympathetic to aspects of the institutional critique should, in her view, consider if they are aiming to create an alternative to effective altruist organizations such as GiveWell; see Rubenstein, 'The Lessons of Effective Altruism', pp. 524–5.

one is moved by the first, the less one can claim to be moved by the second.

Consider, for example, Srinivasan's apparent support for the view that we may have strong reasons to become anti-capitalist revolutionaries (as opposed to effective altruists). This view sits uneasily, at best, with her embrace of the concern that whatever moral response to global poverty we take to be called for does not require that we significantly reduce the time and resources we devote to the values that we care most about (whatever they may be). And this tension is far from unique to her response to effective altruism. Rather, it can be found in a wide range of work in moral and political philosophy, since there is no shortage of thinkers who have hoped to be able to combine ambitious accounts of how our societies and/or the world at large must be changed in order to become just, with moderate accounts of what individuals are obligated to do in response to the overwhelming injustice and suffering that continues to plague our world. The objections to the institutional critique of effective altruism that I have developed here provide, I think, substantial reason to doubt that such a view can be successfully developed. If this is right, then since it cannot be denied that a robust response to global poverty is a moral imperative, we must accept that the appeal of the institutional critique is an illusion.<sup>72</sup>

[bberkey@wharton.upenn.edu](mailto:bberkey@wharton.upenn.edu)

<sup>72</sup> I am grateful to audiences at the 2017 Stanford Junior Scholars Forum on Philanthropy and Civil Society, the Humanitarian Ethics and Action Conference at the University of Birmingham, the Ohio State University Department of Philosophy, the 2016 International Society for Utilitarian Studies Conference, and the 2016 Workshop on Philosophy and Poverty at the University of Salzburg. Paul Brest and Ted Lechterman provided valuable comments at the Stanford forum. I owe special thanks to Eilidh Beaton for very helpful written comments and extensive discussion, and to Lee-Ann Chae for much very helpful discussion. Sarah Light and Amy Sepinwall also provided very helpful written comments and discussion. I have also benefited from discussions with Eamon Aloyo, Ilari Aula, Justin Bernstein, Vikram Bhargava, Constanze Binder, Joseph Bowen, Don Brown, Vince Buccola, Matt Caulfield, Peter Conti-Brown, Chiara Cordelli, Nico Cornell, Monique Deveaux, Linda Eggert, Brad Fulton, Gwen Gordon, Gunter Graf, Elizabeth Harmon, Javier Hidalgo, Joshua Hobbs, Rob Hughes, Corey Katz, Joshua Kissel, Emma Larking, Wesley Longhofer, Amanda Maher, Johanna Mair, Derek Matravers, Tristram McPherson, Kathryn Muyskens, Jonathan Parry, Katarina Pitasse Fragoso, Tony Reeves, Rob Reich, Matthew Rendell, Emma Saunders-Hastings, Gottfried Schweiger, Amy Schiller, Amy Shuster, Peter Singer, Leah Stokes, Alan Strudler, Attila Tanyi, Isaac Taylor, Miles Unterreiner, Makoto Usami, Sara Van Goozen, Ashley Whillans, Richard Yetter Chappell, Carson Young, and Robin Zheng.