

Philosophical Critiques of Effective Altruism

What is most striking in published commentaries on effective altruism written by philosophers is that they are often derisive and contemptuous in tone yet weak in argument. The objections they advance tend to be at least as much *ad hominem* as substantive in character. While I do not find this surprising, I do find it depressing. The primary goal of most of those who identify themselves as effective altruists is the prevention or alleviation of suffering and premature death resulting from poverty and disease in the areas of the world in which these problems are worst, or affect the greatest number of people. To the best of my knowledge, none of the philosophical critics of effective altruism rejects this goal. It is therefore dispiriting to read their criticisms, which often ridicule people who are devoting their lives, often at considerable personal sacrifice, to the achievement of this shared goal, and are often gleeful rather than constructive in their attempts to expose the effective altruists' mistakes in their choices among means.

In this brief article I will discuss some of the criticisms that philosophers have advanced against effective altruism. I will refrain from speculating about the psychology behind the critics' antagonism. The explanations are no doubt complex and various. At the end I will comment briefly on criticisms of effective altruism by developmental economists. The best of these contrast with the philosophical commentaries in being expressed respectfully and in acknowledging that their disagreements are concerned with priorities and with the means of achieving shared ends.

To the extent that the philosophical critics discuss the moral philosophy underlying effective altruism, their criticisms consist almost exclusively of rehearsals of familiar objections to utilitarianism, mainly those presented in the 1970s by Bernard Williams, who was himself more gifted in amusingly ridiculing positions with which he disagreed, and eviscerating their defenders, than anyone else I have ever known. That those who rely almost entirely on his arguments in this context also tend to mimic his polemical style (which in his case appeared more in his conversation than in his publications) may be one unfortunate aspect of his brilliant legacy.

The critics I will discuss tend to assume that effective altruism is grounded in a commitment on the part of its adherents to utilitarianism. That is understandable, as many or even most of those who write and act under that banner are in fact utilitarians whose concern is to produce the greatest good, impartially considered (and whose particular goals may therefore shift depending on which activities promise at a given time to promote the good most effectively). But there is no essential dependence of effective altruism on utilitarianism. Peter Singer's earliest argument in support of a demanding standard for giving to the poor appealed in the first instance to a single widely held moral intuition and argued that consistency required those who accepted

the intuition to give most of their wealth to the relief of extreme poverty. Some years later, Peter Unger, in *Living High and Letting Die*, reasoned in a similar but more systematic way to the same conclusion, explicitly disavowing any commitment to or reliance on a particular moral theory. His aim was to demonstrate that a view of the sort that now informs the work of effective altruists is implicit in values and convictions we already have. Neither his nor Singer's arguments presupposed or implied that we must give equal weight to everyone's well-being. It is therefore insufficient to refute the claims of effective altruism simply to haul out Williams's much debated objections to utilitarianism. To justify their disdain, critics must demonstrate that the positive arguments presented by Singer, Unger, and others, which are independent of any theoretical commitments, are mistaken.

Yet the reliance on Williams alone, without any effort to add to or even to defend his claims, seems irresistible. John Gray, in a review of Peter Singer's recent book defending effective altruism, writes that people's acceptance, at the urging of effective altruism, of what Williams referred to as "negative responsibility" (which asserts that one is responsible for evils that one could have prevented) would be a fundamental compromise of their moral integrity. If this is required by utilitarian ethics, so much the worse for utilitarianism.' (<http://www.nybooks.com/articles/2015/05/21/how-and-how-not-to-be-good/>) Williams thought that the reason that utilitarianism is incompatible with moral integrity is that it alienates people from what may matter most to them: their projects, their relations with those they care about, and so on. It does this because it implies that their projects, commitments, attachments, and values matter no more in themselves, or from what Sidgwick called 'the point of view of the universe', than those of other people. One's own projects and attachments therefore cannot have any priority or privileged role in the determination of how one ought to live or what one ought to do. To accept this, Williams thought, would be to surrender all that makes one's life worth living. There is thus a distinct echo of Williams's own words in Gray's claim that 'for many of us a world in which our own projects and attachments were accorded value only insofar as they enabled us to maximize the general good ... would be hardly worth living in'.

The insistence that morality is essentially first-personal rather than impersonal is an obsessive theme in the writings of the philosophical critics of effective altruism. Gray claims that 'whether or not they find fulfillment in the way they live, effective altruists are bound to view their lives not as ends in themselves but as means to the greatest good'. Amia Srinivasan, reviewing a book promoting effective altruism by William MacAskill, contends that his adoption of an impersonal view, or the point of view of the universe, involves stepping 'outside what is unavoidably the scene of ethical action: one's own point of view'. (<http://www.lrb.co.uk/v37/n18/amia-srinivasan/stop-the-robot-apocalypse>) MacAskill's view thus obscures the fact that you ought to

console your bereaved friend rather than do something else that would be impersonally better, 'not because you've already met your do-gooding quota, but because it's *your* friend that is in distress. This is also the reason you shouldn't deal in subprime mortgages or make money from the exploitation of labour, even if the good effects would outweigh the bad: it's *your* life, and it matters, morally speaking, what you do with it'. Finally, Nakul Krishna, in a reminiscence about his graduate career at Oxford that is also a critique of effective altruism and an encomium to Williams, observes that within utilitarian theory, 'the fact that *I'm me* has been declared, right at the outset, irrelevant' and goes on to assert that there is no irrationality in 'preferring my own point of view to the universe's (whatever that means)'. (<http://thepointmag.com/2016/examined-life/add-your-own-egg> All italics in the originals.)

In short, according to those who criticize effective altruism by appealing to Williams's objections to utilitarianism, the importance to oneself of one's own projects and attachments limits the extent to which morality can demand that one provide assistance to others. But the problem with this claim is that, to the extent that it is plausible, it ought also to apply to other equally or more onerous demands that morality might be supposed to make. If my being me and having my own life can exempt me from the moral reason I might otherwise have to save someone unrelated to me (even though she is *she*, with *her* own life), it seems that these same facts should also exempt me from the moral reason I have not to kill this person *if* killing her were as important to me or my projects as avoiding having to save her is.

In many areas of contemporary moral philosophy (notably population ethics), the disputants are sometimes unaware of previous debates in which some of the points they are making have already been discussed. When this happens, there is inevitably a certain amount of reinventing the wheel – or the flat tire, as the case may be. Some readers may be surprised to learn that less than two decades ago, some of the charges now being urged against effective altruists were pressed against Peter Unger by an unlikely antagonist: Martha Nussbaum. Her lengthy review of his book not only anticipates some of the objections of the critics of effective altruism but also prefigures their polemical style. If anything, it is written in an even more aggressively sneering, dismissive manner, which, at least in my reading of her work, seems quite uncharacteristic.

(<http://www.lrb.co.uk/v19/n17/martha-nussbaum/if-oxfam-ran-the-world>)

Her main argument is intended to be a *reductio ad absurdum* of Unger's conclusion that an affluent person 'must contribute to vitally effective groups, like Oxfam and Unicef, most of the money and property she now has, and most of what comes her way for the foreseeable future'. Her strategy is to invite the reader to 'suppose all the people to whom it is addressed followed Unger's advice: what would the world then be like?' There then follows a detailed vision of a world ruled by Oxfam, which concludes with her comment that 'Unger doesn't even try to imagine this world, and he seems not to have asked himself

any questions about what would actually happen if people took his advice. This would appear to be because he has assumed that people will not take his advice and that he will remain one of a small band of moral heroes, in a world of moral sloth'. In this final comment there is indeed an explanation of why Unger did not consider the dystopia she has sketched. In a reply signed only by Unger but that I coauthored with him (mainly because he did not trust his ability to write a temperate response), we pointed out that 'the injunction she cites was addressed to the conscience of the individual reader in the world as it is ... [It] was, of course, conditional on the wholly realistic assumption that even after my book had its full foreseeable effect, this state of affairs would continue: that for the foreseeable future there would be no radical institutional changes, that most affluent individuals would continue to donate next to nothing, or even nothing at all. ... [P]erhaps some day the world will be receptive to rational reforms of the global economic system. But until this Utopian condition prevails, there is much that a single individual can and should do'.

This is also part of the answer to a similar charge by Srinivasan that effective altruism is 'profoundly individualistic. Its utilitarian calculations presuppose that everyone else will continue to conduct business as usual; the world is a given, in which one can make careful, piecemeal interventions. ... The philosopher is left to theorise only the autonomous man, the world a mere background for his righteous choices'. Although it is presented as an objection, this seems to me exactly right: individuals must decide what to do against the background of what others will in fact do.

Srinivasan goes on, however, to extend the objection, suggesting that the deeper objection is that the effective altruists' focus on individual action prevents them from seeing that the only plausible solutions to problems of global poverty are collective and institutional. 'The tacit assumption' of the effective altruists, she writes,

is that the individual, not the community, class or state, is the proper object of moral theorising. There are benefits to thinking this way. If everything comes down to the marginal individual, then our ethical ambitions can be safely circumscribed; the philosopher is freed from the burden of trying to understand the mess we're in, or of proposing an alternative vision of how things could be.

This claim is echoed by another philosopher, Lisa Herzog, who contends that the institutions of 'the current order' are 'designed and maintained by human beings, and it is up to us, collectively, to reform them. Because of its focus on the "rational choices" of individuals *within* the current system, this is the point that effective altruism misses or ignores'.

<https://www.opendemocracy.net/transformation/lisa-herzog/can-effective-altruism-really-change-world>

Here too both Srinivasan and Herzog have been anticipated by Nussbaum, who complained that Unger 'is basically not interested in institutional and

political issues'. Yet, again, what these philosophers find objectionable is entirely appropriate. I am neither a community nor a state. I can determine only what I will do, not what my community or state will do. I can, of course, decide to concentrate my individual efforts on changing my state's institutions, or indeed on trying to change global economic institutions, though the probability of my making a difference to the lives of badly off individuals may be substantially lower if I adopt this course than if I undertake more direct action, unmediated by the state.

It is obviously better, however, if people do both. Yet there has to be a certain division of moral labor, with some people taking direct action to address the plight of the most impoverished people, while others devote their efforts to bringing about institutional changes through political action. To suppose that the only acceptable option is to work to reform global economic institutions and that it is self-indulgent to make incremental contributions to the amelioration of poverty through individual action is 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.

That some philosophers work to understand what our individual duties might be against a background of malfunctioning institutions does not free 'the philosopher' from trying also to understand issues of global justice and institutional reform. No philosopher I know is looking for reasons to avoid working to achieve an enhanced moral understanding. Yet if others who are not philosophers become persuaded that Srinivasan and Herzog are right that the appropriate agents for addressing problems of global poverty are communities, classes, and states, they are likely to be quite content to leave the problems to those entities and not bother with them themselves.

Each of the three philosophers I have discussed most – Srinivasan, Gray, and Nussbaum – concludes his or her review by claiming that the philosopher whose work they have discussed (MacAskill, Singer, and Unger, respectively) has somehow debased or betrayed philosophy itself. Their closing sentences are as follows. Srinivasan: 'You wouldn't be blamed for hoping that philosophy has more to give'. Gray: 'If history is our guide we can expect Singer's movement for effective altruism to go the way of Comte's church of positivism, which has passed into history as an example of the follies of philosophy'. Nussbaum: 'Philosophy ... offers nothing if not nuance and sustained reflection, and delicate theory-building. In the process of getting philosophy to be more practical, Unger has ultimately sold it out'. Yet both Srinivasan and Gray (though not Nussbaum) write as if the work of moral philosophers is not to be taken seriously in any case. Both repeatedly appeal to sturdy common sense, to what all right-thinking people believe, which they appear to assume is immune to rejection or revision in response to philosophical argument.

Srinivasan assumes that merely to cite MacAskill's views is to expose them to ridicule, and thus introduces her summaries of his views with such

observations as 'the results of all this number-crunching are sometimes satisfyingly counterintuitive'. She is skeptical of the traditional philosophical ideal of following the argument where it leads: 'Philosophers have a tendency to slip from sense into seeming absurdity: a defence of abortion ends up defending infanticide; an argument for vegetarianism turns into a call for the extermination of wild carnivores'. And sometimes it seems a sufficient objection to an idea for which MacAskill may have argued that 'this is not our everyday sense of ethical life'. Our everyday thought, it seems, cannot be overthrown by philosophical argument.

Gray too seems to assume that it is an objection to a conclusion of a philosophical argument that people do not in fact accept it. His central objection to Singer's view is that it 'requires some radical departures from common moral beliefs', and his text is liberally sprinkled with such phrases as 'in the view of most people...', 'in the eyes of many...', 'for most of us...', and so on. In each case what follows the ellipsis is the contrary or contradictory of some view of Singer's. Statements of Singer's view are correspondingly prefaced by phrases such as 'not everyone will share Singer's...', 'nor will many accept that...', and so on. Effective altruism urges people to do the most good they can. Gray observes that 'for most human beings, living ethically is not about doing the most good'. This is true but also largely irrelevant to whether morality in fact demands that we do the most good.

It might seem self-serving for me, a moral philosopher, to express skepticism about the supposition that the truth about morality is already contained in the common sense moral beliefs of ordinary people. For if that were true, the job of substantive moral philosophers (though not of those who work in metaethics) would be merely to fill in some of the details. Nonetheless, my experience as a moral philosopher has been to find that common moral beliefs are often confused and inconsistent (as Singer's and Unger's earlier arguments claimed). Whenever I consider a moral issue with care, I inevitably find the common sense view rather shallow. It is always possible to go deeper. I share the sense, articulated in the final chapter of Parfit's *Reasons and Persons*, that we have only recently begun to understand morality independently of constraints on moral inquiry imposed by deference to ancient religious texts. I therefore think it is a mistake to suppose that the moral views of effective altruists can be rejected on the ground that they are more demanding than people now and in the past have thought that morality could be. It may well turn out that future people will view the failure of affluent people to take individual action to save the lives of people in impoverished areas in much the way we now look back on the drivers of slaves, who were also acting in conformity with 'the view of most people' at the time. It is salutary to recall that the early efforts of those we now recognize as having been in the vanguard of moral progress – abolitionists, campaigners for women's rights and female suffrage, vegetarians and opponents of vivisection –

have always been fiercely resisted and ridiculed by those to whom it was inconceivable that the common sense view at the time might be mistaken.

One may think that I have harped excessively on the fact that the philosophical critics of effective altruism tend to express their objections in a mocking and disdainful manner. But this seems significant to me, as it is suggestive of bad faith. The issues that the effective altruists are addressing are of the utmost seriousness. They should not be occasions for the scoring of debating points or for displays of cleverness, rhetorical prowess, or moral exhibitionism (as when critics, while presenting their objections, pause to reveal parenthetically how much they have donated to charity despite their theoretical misgivings).

Although I have found nothing in the writings of the philosophical critics that has prompted me to alter my view of effective altruism, I have found certain criticisms from economists quite forceful. (See, for example, Emily Clough's "Effective Altruism's Blind Spot" [<https://bostonreview.net/world/emily-clough-effective-altruism-ngos>] and Angus Deaton's *The Great Escape: Health, Wealth, and the Origins of Inequality*.) Developmental economists have, for example, indicated ways in which the efforts of philanthropists, acting through the agency of NGOs, have conflicted with and partly undermined the potentially more effective activities of other agents, particularly representatives of states. And they offer strong evidence that these same activities can retard progress in the long term by helping to keep popular discontent in impoverished societies at containable levels, thereby enabling local dictators to sustain the practices and institutions that keep the population in poverty. Furthermore, by supplying much of what the local government should be providing for its people, foreign aid, whether from NGOs or other states, may enable dictators to use the resources they can gather from domestic sources for the purchase of weapons and the employment of soldiers, again enabling them to resist pressures to change the practices and institutions that perpetuate extreme poverty. While effective altruists can afford to ignore many of their philosophical critics, they should not ignore these other more serious empirical claims.

Jeff McMahan
White's Professor of Moral Philosophy
University of Oxford