Utilitas (2020), **32**, 19–32
doi:10.1017/S0953820819000281

CAMBRIDGE
UNIVERSITY PRESS

ARTICLE

Beyond Moral Efficiency: Effective Altruism and Theorizing about Effectiveness

Federico Zuolo*

University of Genova
*Corresponding author. Email:federico.zuolo@unige.it

(First published online 30 July 2019)

Abstract
In this article I provide a conceptual analysis of an underexplored issue in the debate about effective altruism: its theory of effectiveness. First, I distinguish effectiveness from efficiency and claim that effective altruism understands effectiveness through the lens of efficiency. Then, I discuss the limitations of this approach in particular with respect to the charge that it is incapable of supporting structural change. Finally, I propose an expansion of the notion of effectiveness of effective altruism by referring to the debate in political philosophy about realism and the practical challenge of normative theories. I argue that effective altruism, both as a social movement and as a conceptual paradigm, would benefit from clarifying its ideal, taking into account the role of institutions, and expanding its idea of feasibility.

I. Effective altruism: theory of a practice

Effective altruism (henceforth EA) is one of those few intellectual positions where philosophical theorizing has directly given rise to a powerful social movement. The striking peculiarity of EA is its commitment to connect one of our most general, and yet vague, duties – namely that of doing good – with some practical, and possibly really effective, envisaged outcomes. What is unique to EA is its promise of uncompromising ethical clarity brought about in a practically effective manner.

Many have discussed the moral premises of EA. In this article, instead, I provide a critical analysis of an underexplored aspect of EA, namely its effectiveness dimension. I will thus not address other fundamental issues regarding the ethics of giving, such as how much and why we ought to donate, to whom we should donate first, whether there is a conflict between the duty of impartiality and legitimate partial commitments, the altruistic repugnant conclusion, and so on.[1] In particular, I will try to understand

[1] W. MacAskill, *Doing Good Better: Effective Altruism and How You Can Make a Difference* (London, 2016); J. McMahan, ‘Philosophical Critiques of Effective Altruism’, *The Philosophers’ Magazine* 73 (2016): pp. 92–9; T. Pummer, ‘Whether and Where to Give’, *Philosophy and Public Affairs* 44 (2016), pp. 77–95; P. Singer, *The Most Good You Can Do: How Effective Altruism is Changing Ideas About Living Ethically* (New Haven and London, 2015); G. Pellegrino, ‘Effective Altruism and the Altruistic Repugnant Conclusion’, *Essays in Philosophy* 18 (2017), pp. 1–24.

whether the conception of effectiveness that EA embeds is coherent with the underlying assumption that our duty is to bring about as much good as possible.

The article will proceed as follows. In the next section I provide a definition of effectiveness and try to understand better EA's conception of effectiveness. In the third section, I distinguish effectiveness and efficiency and shed light on EA's penchant for efficiency. Then, I briefly mention the main critiques offered against EA. In the fifth section, I highlight other notions related to effectiveness that have been discussed in debates about feasibility, realism, and the methodology of political philosophy. Building on these notions, in the sixth section, I seek to expand EA's theory of effectiveness so as to understand how EA can respond to some of the critiques outlined. The final section provides a sort of agenda for effective altruists (henceforth EAs) to make EA more in line with a conceptually and practically deeper understanding of the commitment to effectiveness.

II. A definition of effectiveness

Given its theoretical formulation and ambitions, I shall now analyse EA's theory of effectiveness from a theoretical point of view, namely not by checking whether the actions recommended by EAs are actually effective in practice. But what is effectiveness about? In a proper sense, only an action (or a set thereof) can be effective. Of course, theories are primarily effective on people's thoughts and, as a consequence of this, on actions. However, given EA's insistence on providing practical recommendations, I will focus here only on actions. Hence, I will refer to the effectiveness of the actions that the theory recommends were they put into practice.

> An action (A) is effective (E) with respect to a value (V) iff the state of affairs (SoA) brought about by (A) embeds (V).

This definition is binary because it rests on the *ex post* certainty of knowing what has in reality been effective. It can also be formulated in a scalar sense by asking how much (V) an (A) can actually bring about. A scalar definition is particularly suitable when we need to be probabilistic, that is, when we need to estimate in advance the expected effectiveness of an action in a condition of uncertainty.

> The expected effectiveness (EE) of an action regarding a value (V) can be measured with the probability (P) that a (SoA), which embeds (V), is reached given the action (A), which implements (SoA).

Building on this, EA's account of effectiveness can be put as an answer to the following question: how could we maximize the expected outcome given our limited epistemic and practical resources? It seems rational to suppose that given our limited resources, the complexity of the world and the urgency of some problems, we would do better to focus on problems that we can expect can be solved by agents about whom we have reliable information. Three features seem obvious candidates to make sense of this.

The first two features are *directness* and *specificity of goals*. I define a goal as the condition in which a desired (V) is embedded in a (SoA). The logic behind EA implies that in order to be effective, we must preferably aim at addressing specific goals that we can expect to reach through a foreseeable course of actions made by recognizable agents. This formulation seems reasonable enough. After all, who would want to engage

https://doi.org/10.1017/S0953820819000281 Published online by Cambridge University Press

with the realization of vague goals for whose implementation we have no clear route? The third feature concerns the *measurability* of effectiveness. On EA's view, in order to meet criteria of rationality, the expected effectiveness of an (A) should be measurable.

These three features are related to a further overall trait of EA: its attempt to reduce complexity. Usually, acting and making such an action effective in bringing about the desired state of affairs may be extremely complicated. We may fail to bring about good states of affairs for many obvious but not fully predictable reasons. To make a long story short, there are at least objective features (the world is complex and the causes of injustice are intricate), subjective features (people are often selfish or irrational in their behaviour), and intersubjective features (people have competing legitimate interests).

With respect to these features, EA provides reasons to reduce the problems stemming from the subjective area, by demanding that people focus on what one is capable of doing altruistically (typically donating). Moreover, it *de facto* ignores the intersubjective issues because it assumes that actions to address urgent and real priorities – such as the reduction of extreme poverty or fatal diseases – do not conflict with each other. And, finally, it has a very peculiar strategy to (not) deal with the objective complexity of the world because it usually recommends simply donating to the most effective charities or NGOs, whose effectiveness has been assessed by meta-charities.

In sum, EA's theory of effectiveness may be summarized with the following points:

> Not any goal is compatible with an effectiveness-driven approach. Only measurable goals are. How much (V) is included in a (SoA) should be measurable.
>
> We should bring about the (SoA) that maximizes the amount of (V).
>
> In theory, we can devise only an action (or a restricted set thereof) that brings about such a (SoA) maximizing the desired (V).

III. Effectiveness and efficiency

So far so good. These features seem to flow from some general prescriptions of rationality applied to hard choices in a realistic scenario. In what follows I shall argue that, unlike what some EAs have maintained, EA is, in fact, not about effectiveness generally conceived. Rather, it concerns a specific way of pursuing effectiveness, namely *efficiency*. What is the difference between effectiveness and efficiency? Although they are sometimes used interchangeably, we can safely say that they differ in the following way. Effectiveness concerns the overall capacity of an action actually to reach a goal, while efficiency concerns whether a certain goal is reached with the best use of resources. We may say, for instance, that a procedure or technology is more efficient than another because it needs fewer resources to generate the desired output. Or we may say that a strategy was effective but wasteful, namely non-efficient.

I define efficiency in the following way.

> (A) is efficient to the extent that, given a fixed amount of resources (R), it best employs (R) to reach (V) in (SoA) with respect to other feasible alternatives.

Hence, efficiency, unlike effectiveness, is dependent upon a parametrical amount of resources. Moreover, it can only be a scalar notion and its function is mostly comparative in adjudicating between the competing alternatives.

https://doi.org/10.1017/S0953820819000281 Published online by Cambridge University Press

I say that EAs seem mostly concerned with efficiency for the following reasons. First, they repeatedly aim at calculating the impact that a unit of input (say $1) can have in terms of units of goodness – say numbers of lives saved, or impact in terms of Disability-adjusted life year (DALY) and Quality-adjusted life year (QALY). Many EAs, and in particular Pummer and MacAskill, argue for the conditional duty to be efficient upon deciding to donate. In other words, they claim that we have a duty to generate the maximum expected amount of output, given a certain amount of resources (typically a portion of one's income). Moreover, MacAskill repeatedly mentions the criterion of cost-effectiveness, which is efficiency in other terms.[2]

Second, they seem to urge us to be concerned with both internal and external efficiency in the use of resources. Internally we ought to choose our career by aiming at the one whose expected salary is the highest so that we can donate more. EA also aims at external efficiency because charities are to be selected according to their efficiency in getting most good (measured, for instance, in terms of lives saved) per monetary input received.

Third, they recommend that we assess the value of our actions – regarding donation but also the choice of our career – given the law of diminishing marginal returns, namely in respect of how much good our action or career can achieve given the fact that our action and career is only a marginal addition to the actions and work of others.[3]

It is unclear to me how much this conflation of effectiveness with efficiency stems from a lack of understanding of the difference between the two concepts, or from a rhetorical emphasis that EAs put in using the idea of effectiveness in order to make it sound more attractive. To my knowledge, only Pallotta has argued against the idea that EA is about efficiency.[4] But he takes too narrow an understanding of what efficiency means. By efficiency he means the organization or procedure that costs less, the so-called 'admin-to-program ratio', which does not take into account what is actually achieved in practice by the organization or procedure. My argument is immune to this reply because the definition of efficiency I employ does not collapse into the mere minimization of costs and like the definition of effectiveness includes the notion of goal, which is what counts in practical matters. If this interpretation is convincing, it shouldn't be seen as a disappointment. After all, as noted, efficiency is a prominent way of contributing to effectiveness. But it is not the same thing, given its specific focus on how to reach the outcome given a limited amount of resources. Of course, I do not want to argue that pursuing efficient interventions is wrong and that EAs should aim at inefficient strategies. Rather, the question is whether this understanding of effectiveness via efficiency lives up to its ambitions.

Critics point out that one cannot really aim at improving the condition of the worst-off without tackling the structural causes of poverty, that is, without addressing institutional issues of injustice, oppression, the political dimension of disadvantage, etc.[5] Brian

[2] MacAskill, *Doing Good Better*, p. 135.

[3] See in particular MacAskill, *Doing Good Better*.

[4] D. Pallotta, '"Efficiency" Measures Miss the Point', *The Effective Altruism Handbook*, ed. R. Carey (Oxford, 2015), pp. 32–4.

[5] On these critiques, see Lisa Herzog, 'Can "Effective Altruism" Really Change the World?', *Open Democracy*, 22 February 2016, <https://www.opendemocracy.net/transformation/lisa-herzog/can-effective-altruism-really-change-world>; and A. Srinivasan, 'Stop the Robot Apocalypse', *London Review of Books* 37 (2015), pp. 3–6. For a more detailed analysis of why EA is not equipped to understand the importance

https://doi.org/10.1017/S0953820819000281 Published online by Cambridge University Press

Berkey defines this argument as the 'institutionalist critique'.[6] By refusing to address these issues, EAs would fail to tackle the real underlying causes of the most pressing problems that EAs are committed to solving. Moreover, some have argued that EA's specific recommendations might have perverse effects in creating alternative providers of services (for instance education) that would be typically used by those who are not the worst-off, thus leaving the worst-off dependent upon the state service whose quality is likely to be diminished because of the lower pressure that people put on state agencies.[7] Such effects are hardly detected by *ex post* assessments of aid, whose perspective is usually short-term and extremely targeted, thus overlooking the possibly perverse or unintended effects of aid.[8] In sum, such critics claim that EA lacks insight into the structural causes of moral wrongs because EA is concerned only with the effects of deep problems. Hence, the critical argument concludes, in overlooking the systemic and long-term issues, EA does not live up to its practical commitments.

In many cases these remarks seem plausible. But to be fair to EA, EAs also have a point in focusing only on actions that can actually bring about real and measurable, albeit 'minor', improvements. In one sense this approach is likely to be very effective, at least if we compare it to other available alternatives. As Berkey argues, EAs may have mistakenly assessed the probabilities of enacting change, but, unlike others, they 'have at least attempted to engage with these challenging issues'.[9] In general, EAs seem to rely on the assumption that we should practically care only about issues on which we can have a sizeable and expectedly direct impact. To be sure, changing the background condition of society might in the end be much more effective than specific micro-interventions. But how can we be sure of having an impact in the long run and in relation to such complicated issues?

From the point of view of political theory, EAs seem both hyper-realists and moralists. On the one hand, they are in a sense hyper-realists in that they take the global order and the institutional setting almost for granted and not as possible targets of change. This is so because they think it is beyond our foreseeable control. On the other hand, in so doing they recommend doing as much as we can, thus being very demanding moralists.

We have seen that EAs are committed to a sort of efficient management of the resources that have to be devoted to donation in order to yield as much good as possible. However, this focus on efficiency is not necessarily the best way to pursue effectiveness in general. Of course, an efficient action is *also* effective because efficiency presupposes effectiveness and I am not endorsing the adoption of inefficient solutions. Rather, the point I want to make is that efficiency is a specific way of understanding effectiveness that may be misleading in some cases. This is especially evident if we consider that there are many relevant causes whose advancement requires and has required actions that we would not necessarily qualify as efficient. For instance, Gandhi's pacifist movement for India's independence, M. L. King's march for the rights of African-Americans, and Mandela's fight for the end of the apartheid regime were all complex

of structural change, see T. Syme, 'Charity vs. Revolution: Effective Altruism and the Systemic Change Objection', *Ethical Theory and Moral Practice* (2019), pp. 1–28, doi: 10.1007/s10677-019-09979-5.

[6] B. Berkey, 'The Institutional Critique of Effective Altruism', *Utilitas* 30 (2018), pp. 143–71.

[7] On this see E. Clough, 'Effective Altruism's Political Blind Spot', *Boston Review* 14 July (2015).

[8] On this see L. Wenar, 'Poverty is No Pond: Challenges for the Affluent', *Giving Well: The Ethics of Philanthropy*, ed. P. Illingworth, T. Pogge and L. Wenar (Oxford, 2010), pp. 104–32.

[9] Berkey, 'The Institutional Critique', p. 162.

https://doi.org/10.1017/S0953820819000281 Published online by Cambridge University Press

sets of activities in which the leaders and ordinary people put their life at risk without complying with an efficient use of their personal resources. We know now that these social movements were eventually effective because the goals were reached. And we know that many people supporting them, in particular their leaders, relied on a strong faith in the sense of their action. Such faith entailed both a hope of securing the final success and a staunch commitment to the morality of their cause, whatever the probability of success. However, if we do not consider the *ex post* success of these movements, how could we assess these expectations were we to find ourselves at the beginning of their activism? Besides the obvious duty that one has to do the right thing, were these actions rational in terms of their expected probability of success? It does not seem an exaggeration to say that the probability of success should not have been considered high because the initiatives were also very complex and involved a host of intermediate and indirect actions in order to reach the goal. And, even if one thought that in the end the cause would have been won, were the actions that the movements and leaders undertook acceptable (or the best) in terms of efficiency? To be sure, one may say that in calculating whether these activists should have engaged in their actions, we should factor in not only the very low probability of success, but also and most importantly the very high amount of possible gains, thus making the case for the moral necessity to take these initiatives. That is a sensible remark and it would perfectly fit in with the spirit of EA. Indeed, depending on the ratio between the assessment of the probabilities of success and the amount of possible gain, EA may justify either a rather piecemeal and short-term set of engagements, or a long-term and radical set of initiatives.

All these questions are of course very controversial and difficult to answer. What I want to point out, though, is that the logic of EA would have hardly *ex ante* justified these actions. But these actions were eventually effective and extremely beneficial to billions of people. Hence, what should we think about EA's theory of effectiveness? Shouldn't we revise it in order to take into account these historical examples?

In the interests of fairness, we should note that MacAskill claims that EA is not reducible to charitable donations because EA can also justify and recommend actions that are not easily quantifiable if the expected gains are high (as in the case of political careers).[10] However, in making reference to a political career and other non-easily quantifiable expected goals, whatever MacAskill actually thinks about the desirability of these options, he does not spend much time discussing them because the conditions that justify engaging in a political career are rarely met. As Iason Gabriel convincingly points out, EA is wedded to the logic of frequently changing the priorities and addressees of donations in virtue of their varying marginal effectiveness.[11] Hence, it seems ill-suited to the pursuit of goals requiring long-standing, stubborn, and patient activism, which might only eventually be effective.[12] In theory, supporting structural change can be justified in EA's framework. However, it is unclear how EAs could choose to pursue a long and structural cause over the achievement of a specific but substantial improvement. At what point in the ratio between probability of success and magnitude of

[10] MacAskill, *Doing Good Better*, pp. 114–21.

[11] I. Gabriel, 'Effective Altruism and its Critics', *Journal of Applied Philosophy* 34 (2017), pp. 457–73.

[12] The commitment to marginal effectiveness also has the paradoxical effect of suggesting that people abandon a previously 'underdog' cause once it becomes popular and begins to have major effects. See J. Kissel, 'Effective Altruism and Anti-Capitalism: An Attempt at Reconciliation', *Essays in Philosophy* 18 (2017), pp. 1–23, at 19.

https://doi.org/10.1017/S0953820819000281 Published online by Cambridge University Press

the expected achievement should we opt for the initiative with low probabilities of success but which tackles a structural cause? If, as EAs claim, the choice is always comparative, the fate of structural causes seems doomed in so far as one can always find a more urgent and specific cause which we should choose in virtue of its achievability. One may reply that EA prefers measurable and specific goals to wider social changes because there is no sufficient good evidence of the preferability of the latter. Hence, EAs are not in principle against structural change for they only want their actions to make a real contribution and not be in vain. This point makes sense, but it gives further support to my overall claim that, even though EA could in principle support any cause, it *de facto* cannot but opt for certain causes.

But, one may retort, we should not forget that EAs' reflections are not limited to minor, although important, improvements to the status quo, for they are also concerned with the existential risk that events like asteroids striking Earth might pose to human life. This demonstrates that EA's methodology is not just suited to charity, for it may also undertake highly speculative analyses. However, it seems very strange that EAs typically focus on either very close and concrete goals or very worrisome but distant and speculative ones. This 'cross-eyed' perspective leaves out what is in between these two extremes: social issues. Although EAs argue that there is nothing in principle that prevents EA from tackling structural social issues, practical urgency and some methodological biases probably discourage EAs from properly considering them.

Addressing structural causes, indeed, is unlikely to be the result of EA's methodology because we should have clear data on the impact of an individual's contribution and contributing to such a cause should have the highest marginal return compared to other actions to address other problems. But how can an individual action meet this requirement given that in structural causes individual contributions may have such a return only if the cause reaches a tipping point after years of relentless but invisible efforts?

In view of these theoretical difficulties, and the fact that at least some major social forms of activism for the good do not seem to follow EA's theory of effectiveness as efficiency, shouldn't we revise it?

If, instead, EAs would prefer to stick with their theory in which efficiency, measurability, and directness play a paramount role, they should seek an alternative justification for this account, which does not rely solely on its effectiveness-value. Paradoxically as it may seem, a Stoic interpretation of EA's focus on the internal motivation to donate might be the best way to defend EA. Recall the famous Stoic maxim as expressed by Epictetus: we ought to be concerned only with what depends on us. What does not depend on us should not be a matter of moral and existential concern. So, in this view, what depends on us? Our reactions to external events and our motivation to undertake actions. The accomplishment of such actions, or the causes of events that perturb our inner disposition do not depend on us.[13] From this it follows that our control should be directed only towards our inner dispositions, not towards what happens in the world, which is beyond our control. Of course, the Stoic's concern with one's inner sphere does not mean retreating from the world. Rather, it simply concerns how one should react to what happens in the world.

Why do I find these considerations similar to EA's recommendations? Paradoxically, like the ancient Stoic precepts, EA's recommendations similarly demand that we primarily focus on those things that we can control directly. Of course, unlike the Stoic

[13] Epictetus, *Handbook of Epictetus* (Indianapolis, 1983), §1: 11.

https://doi.org/10.1017/S0953820819000281 Published online by Cambridge University Press

principle, EA's principle of efficient moral concern does not say that we should not care about what is not in our control. If we did so, we should not care about the overall amount of poverty, and so on. However, EA's principle of efficient concern may be interpreted as holding that one should not be *practically concerned with and try to achieve what is not under one's control*. What is under one's control is, first, one's motivation to help and donate, and, second, the direction of this motivation, namely the kind of good cause that one can choose. How to achieve the good in practice, and in particular how best to achieve it, is not under one's control and hence should be, so to speak, 'externalized' to meta-charities and NGOs. This Stoic interpretation seems suggestive and may be further explored, but I do not want to overstate the case here.

To conclude, whatever the best interpretation of EA's effectiveness, EA has a peculiar strategy for dealing with the complexity of the world. From a practical viewpoint, EA operates a sort of double externalization: the epistemic difficulty in assessing who's effective is outsourced to meta-charities, and the pragmatic difficulty in devising the most effective action and implementing it is externalized to the charities and NGOs. In sum, EA provides a hyper-control of one's internal disposition and motivation towards donating, while it externalizes the understanding and implementation of actual courses of action because it recommends an efficient management of diverse activities according to one's most high-yielding capacities.

IV. Effectiveness and other cognate notions

EA's account of effectiveness seems reductive in so far as it sees effectiveness as chiefly a matter of efficiency, and it is ill-equipped to drive structural social change. But is this necessarily the case? Can we broaden EA's account of effectiveness so as to address these challenges? To this end, EA's theory of effectiveness should be put in a broader network including other cognate notions that are relevant for the practicality of normative theories. In particular, in what follows I will briefly present some ideas laid out in the debate about the concept of feasibility and realism of political theory. It will not be a complete list of all relevant questions. Rather, it should be understood as a map of the kinds of cognate notions that a theory of effectiveness should take into account.

First, within this debate many meta-theoretical questions have been aired. Some have wondered what should be the appropriate level of fact-sensitivity for a theory of justice. Following G. A. Cohen's critique of Rawls, the issue of sensitivity to facts has been understood in terms of whether and to what extent principles of justice, in order to be valid and sound, should depend on (social or natural) facts. Different answers have been proposed, ranging from those holding that first principles of justice should be independent of facts,[14] to those arguing that facts are relevant not only to the implementation of principles but also to their definition because, among other things, the fact-insensitive level includes underdetermined principles, with little practical guidance.[15]

Related to this question is the distinction between ideal and non-ideal theory. This distinction has been phrased in two connected but diverse ways. The Rawlsian formulation understands ideal theory as the assumption according to which, under favourable conditions, the parties comply with the principles of the theory. This assumption is

[14] G. A. Cohen, *Rescuing Justice and Equality* (Cambridge, MA and London, 2008).

[15] E. Rossi, 'Facts, Principles and (Real) Politics', *Ethical Theory and Moral Practice* 19 (2016), pp. 505–20.

https://doi.org/10.1017/S0953820819000281 Published online by Cambridge University Press

meant to represent how a society would fare if all the parties were to follow the principles of the theory so as to show what the duties in these conditions are.[16] Accordingly, one may ask what the duties are, given a situation of non-compliance (non-ideality) of the parties. The other sense of the distinction between ideal and non-ideal theory is more in line with the issue of fact-sensitivity just noted. In the recent debate on the shape that a normative political theory should have, many have outlined the distinction between ideal and non-ideal theory as a continuum from the least fact-sensitive, where highly abstract principles are presented and defended, to the theory that includes and mirrors a number of empirical and social facts. Hence, the question concerns what level of fact-sensitivity is appropriate to discharge the functions of a normative political theory in terms of its capacity both to guide actions and to be justified independently of the varying contexts.

Second, effectiveness is related to feasibility. By definition, what is effective is also feasible. Feasibility is a modal concept that concerns whether a state of affairs can be brought about by individual or collective actions. As conceptualized by Pablo Gilabert and Holly Lawford-Smith, (political) feasibility can be understood as both a binary concept and a scalar concept.[17] To establish whether a state of affairs is binary-feasible we should assess whether the action to bring about the state of affairs is compatible with some hard constraints, established by the laws of logic, physics, and biology. Scalar feasibility can be measured, instead, in terms of the probability of an action resulting in the desired state of affairs, given some soft constraints (cultural, economic, social). The types of feasibility have different effects on our duties and practical recommendations. Normative principles or political proposals that violate hard constraints are unfeasible – thus violating the 'ought-implies-can' principle – and should not be considered valid, whereas the different scores of diverse principles and proposals in the scale of scalar feasibility should be weighed in order to evaluate the preferability of practical alternatives.

Third, a further distinction that is relevant here is that between the access-dimension of feasibility and the stability-dimension of feasibility.[18] Access-feasibility concerns the route that might lead from the current state of affairs to the desired state of affairs. This requires asking which agents and actions could bring it about and under what conditions. Stability-feasibility concerns whether the just state of affairs can actually be stable over time. Stability depends mostly on the compatibility between the demands of the institutions and rules embedded in the desired state of affairs and the human motivational makeup. It also concerns the issue of whether the actions required by the normative theory may be, first, immune to perverse effects or self-effacement, and, second, self-sustaining, that is, capable of generating autonomous motivation.

The final issue that is pertinent to our practical commitments towards a just world is that of reconciliation. Rawls claimed that one of the tasks of a normative theory is that of reconciling ourselves with social reality, by showing that current rules and institutions and their history display a rational and justifiable form.[19] The importance of reconciliation is that it shows that what we seek through normative theory is at least

[16] J. Rawls, *A Theory of Justice*, rev. edn. (Cambridge, MA, 1999).

[17] P. Gilabert and H. Lawford-Smith, 'Political Feasibility: A Conceptual Exploration', *Political Studies* 60 (2012), pp. 809–25.

[18] H. Lawford-Smith, 'Understanding Political Feasibility', *The Journal of Political Philosophy* 21 (2013), pp. 243–59.

[19] J. Rawls, *Justice as Fairness: A Restatement* (Cambridge, MA, 2001), p. 3.

https://doi.org/10.1017/S0953820819000281 Published online by Cambridge University Press

partially present in reality, thus pointing out that reality is not completely immoral or irrational, and that our endeavours to realize justice more fully are not doomed or inane.

V. Expanding EA's conception of effectiveness

With respect to all these questions, EA's overall relation to practice seems original but sketchy. Indeed, most of the questions posed in the previous section are only implicitly answered in EA's writings. In what follows, I will seek to tweak EA's account of effectiveness with a view to answering these questions and proceed a little further.

First, regarding fact-sensitivity of first principles and the level of ideality of the theory, EAs take first principles (e.g. that suffering and poverty are bad and that we should alleviate them) to be self-evident, unquestionable and accessible to all well-meaning people. Moreover, they are thought to be valid and compelling in any condition, be it ideal or fully non-ideal. However, the practical and theoretical approach of EA is set up in a markedly non-ideal fashion. This might be a problem because whether and to what extent moral principles are binding and action-guiding also depends on the favourability of conditions and on the compliance of other people. For instance, the moral principle of helping others actually translates into a duty to donate only depending on whether people are in need and the extent to which other people are also complying with this principle. A world of altruistic donors would not only be impossible, *qua* too demanding, but also *qua* practically ineffective in so far as the duty to donate is conditional upon the existence of needy people, the availability of resources and the lack of other channels to meet the interests of needy people.

Regarding feasibility, EAs typically recommend reliance on the effectiveness-assessment of meta-charities that evaluate the reliability of NGOs and charities. However, this *ex post* approach may be biased. By only relying on safe and confirmed cases of success, EA diminishes the risks of venturing into actions whose capacity to reach the goal is uncertain, but it also limits itself to what we already know and what we have already done. Of course, EA's principles might be applied to explore the effectiveness of untried methods. However, it is unclear to me what conceptual and methodological resources should be employed to assess these methods.

In a sense, EA's conception of effectiveness is surprising and paradoxical. On the one hand, it is rather 'moralistic' in that it targets individuals' sets of motivations and attitudes requiring that each devote a significant portion of their income to donation. On the other hand, it may seem conservative because, as we've seen, tackling structural issues, while not impossible, may only be justified under rare circumstances. In sum, this overall conception of effectiveness is demanding but non-ideal, and focuses on the most pressing problems without being fully concerned with the causes of these problems. It is neither a radical view, nor a conservative one. In short, it is a sort of *remedial effectiveness*, in so far as it is only concerned with the salient expressions of problems without addressing the underlying structures thereof. Building on this, we may seek to outline what EA could say on the other issues related to effectiveness even though they have not been outlined yet.

First, one may ask: what could EA say about access-feasibility? The answer to this question depends on how we understand the ideal or end-state that EA is seeking to advance. On the one hand, we might interpret it as access to the condition where more (or most) people donate and behave altruistically. Hence the remedy to the most pressing moral problems is put in place by altruistic actions. On the other hand, if we think that altruistic actions are not the solution but mere proxies to a better

https://doi.org/10.1017/S0953820819000281 Published online by Cambridge University Press

state of affairs, we should ask what this better state of affairs consists in. On this we have no clear indicators, as EAs seem to rely on the intuitive idea that certain moral wrongs are clear enough, and that we ought to convince people to act in order to solve these problems, rather than convince people why these situations are moral problems. Such an end-state would be a condition where most pressing moral problems are solved. But are we sure that it is uncontroversial to establish how a society without these problems would look? This is hardly so, and EAs would be better having at least a working idea of what kind of end-state they are working towards.

If we suppose that EAs are not completely happy with the structure of current states of affairs and rather think that some institutions ought to be changed, what kind of transformative effectiveness would be compatible with the principles of EA? The kind of changes required will depend on other issues. People like Singer would now consider it unnecessary to change the capitalist system because it has improved the conditions of many people. Some other activists and commentators have held that EA actually 'loves systemic change'.[20] However, it is not clear what this might mean – a more democratic society, a more affluent one, or a vegan one. EA is currently mostly concerned with charity but perhaps it is not necessarily wedded to this. To avoid the objection that it cannot support structural change, perhaps EAs should clarify the nature of EA. Either EA is a moral method to solve pressing problems that are already established as such, or EA is a substantive specific moral theory: in both cases, a specification of EA's ideal is lacking. If EAs opt for the former alternative, then EA should rely on people's intuition about what are the most pressing causes and we would still not have a definition of EA's ideal. If EAs opt for the latter alternative, they can better vindicate EA's theoretical and methodological strength, but they could no longer present EA as a neutral practical method.

Next let us turn to the stability question. On the one hand, this question depends on whether EA demands a transformation of our current world – a transformation which, as we have just seen, remains unspecified. On the other hand, we can try to provide an answer even if there is no transformative concern in EA. We can ask what the world would look like if all or at least a conspicuous number of people followed EA's prescriptions, even within the current institutional system. David Schmidtz has argued that if people were to strictly follow the ethics of giving – which, for our purposes, is the same as EA – the current world would actually collapse because people would stop spending money on a vast array of issues that would be considered futile.[21] But if so, large portions of our economic system would simply be destroyed. This is clearly undesirable from many normative perspectives, including the ones typically endorsed by EAs. A supporter of EA may retort that Schmidtz's conclusion is unwarranted because the application of EA's principle by the majority of people would not necessarily lead to economic collapse. Whatever conclusion is the correct one, in theories of EA there seems to be something missing regarding the relation between the ideal, if any, and its application in reality. Is EA a universal account that applies to any realistic domain, or is it applicable only under (our) very non-ideal conditions?

What if we could convince most people to become EAs? What would a world embedding the principles of EA look like? To make it more functioning than in Schmidtz's imaginary world, we should also consider some institutional issues.

[20] <https://80000hours.org/2015/07/effective-altruists-love-systemic-change/>.

[21] D. Schmidtz, 'Islands in a Sea of Obligation: Limits of the Duty to Rescue', *Law and Philosophy* 19 (2000), pp. 683–705.

Hence, we should imagine what institutional rules there could be, if they were to be underpinned by EA. We might think of it as a society in which all or most people are moved by altruistic motivations, thus retaining only a portion of their work for themselves. Such a society would not generate an egalitarian sharing of the products of social cooperation, but it would certainly make substantive redistributions through acts of voluntary giving. However, that would still amount to an ethical ideal with little institutional framework. To tease out its features, it is perhaps helpful to compare it with another quite moralistic theory that similarly gives a prominent role to altruistic motivations: G. A. Cohen's idea of an egalitarian ethos. Against Rawls, Cohen famously claimed that an egalitarian ethos is a necessary component of a just society and a

> society is more just when its members do not require inequality-generating incentives but rather, inspired by distributive justice (and subject to a personal prerogative), consider the interests of others when making productive choices (both about how many hours to work and what career to pursue).[22]

In Cohen's just society, morality demands that one forsake market-driven (monetary) incentives to be mostly productive and contribute to social cooperation. In this society, people should accept being paid less than their real contribution to social cooperation, in virtue of people's commitment to the egalitarian ethos. EAs take the opposite route and claim that market-driven incentives (and institutions) should be employed to maximize social production, so that altruistic individuals may donate larger shares of their income to the neediest. It is unclear, though, whether such a motivation can and should remain purely a matter of individual ethical behaviour in a framework where the institutions not only do not reinforce it but also cause some of the problems that altruism addresses.

Finally, what could EA say about reconciliation? In one sense, EA is committed to providing grounds for reconciliation. In giving people a moral motivation to abide by current rules and practices while aiming at doing as much good as one can, EA has the (unintended?) effect of reconciling people with their jobs and social functions, while requiring an important change in the personal advantage that one takes from one's position. Perhaps a kind of reason for reconciliation with the economic structures can be found in Singer's words, where he says that, irrespective of whether we value equality as an intrinsic or extrinsic value, capitalism has certainly increased inequalities but also 'lifted hundreds of millions out of extreme poverty'.[23] Hence, the worries about being complicit in a rotten system should be weakened because capitalism's overall consequences are positive.

VI. Conclusion

In conclusion, let me sketch out some considerations on what EAs could do to complete their account of effectiveness. If what I have argued is correct, EAs interpret effectiveness in terms of efficiency of individual contributions. I don't claim that this perspective is wrong. Indeed, it may be interpreted as a way to guarantee that one's contribution

[22] P. Tomlin, 'Survey Article: Internal Doubts about Cohen's Rescue of Justice', *The Journal of Political Philosophy* 18 (2010), pp. 228–47, at 231.

[23] Singer, *The Most Good*, p. 50. It is worth remarking that this positive evaluation of capitalism is conditional and that Singer was previously more critical of it.

https://doi.org/10.1017/S0953820819000281 Published online by Cambridge University Press

does bring about positive consequences. There seems to be a commonsensical concern to avoid wasting resources and people's effort. However, this perspective seems limited because: (i) it cannot *de facto* track structural social changes; and (ii) it is limited to an *ex post* assessment of effectiveness on the basis of what has been effective, thus restricting the field of possibilities.

These two features are understandable in light of the very limited epistemic and practical resources that each of us has. As said, on this I partially agree with Berkey's defence of EA against the institutionalist critique. By contrast, I am less convinced by Jeff McMahan's argument against the same kind of criticism:

> 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 impoverished people 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.[24]

I agree that a certain division of moral labour is inevitable. However, from this the restriction at the beginning of the quoted passage does not follow. It is true that, strictly speaking, I can only determine what is under my control, namely what 'I will do' because I cannot determine 'what my community or state will do'. However, I can demand that my community or state take a certain initiative. Although influencing the state is not under my control, I have many ways of exerting some pressure (voting, public campaigning, demonstrating, striking, engaging in civil disobedience, and so on). Moreover, what is morally relevant here is not only the causal capacity of agents to determine what they and others can do. Having legitimate expectations and advancing moral claims on institutions is a fundamental part of the moral interplay between individuals and collective institutions. What we demand of states and institutions does not automatically and directly translate into the state's initiatives, but it is not a waste of time or a form of exoneration from one's responsibility towards those who are in serious need. This is particularly the case if the causes of this need are dependent upon and enforced by an institutional system.

In sum, it is understandable only to focus on donations and NGOs if we exclusively take the point of view of what individuals *qua* separate individuals ought to do. However, we are members of moral communities which consist of a web of mutual obligations and corresponding reciprocal expectations.[25] Hence, as a sort of constructive proposal, I claim that EAs should take into account the possibility of adopting a perspective that includes public institutions too, both as actors and as partners of initiatives in EA.

How, then, should the theory of effectiveness change? Besides not restricting itself to an efficiency-driven approach, it should perhaps not only focus on individuals' choices. Without dismissing the importance of individual and private initiatives, durable and fundamental change cannot be brought about without institutions and states. How

[24] McMahan, 'Philosophical Critiques', p. 95.

[25] In addition to individual obligations, people also have collective obligations. On this see A. Dietz, 'Effective Altruism and Collective Obligations', *Utilitas* 31 (2019), pp. 106–15.

https://doi.org/10.1017/S0953820819000281 Published online by Cambridge University Press

can EAs estimate the probability of success of such agents? That is impossible to say in general. Perhaps only local solutions may be found. However, other variables may be taken into account regarding, for instance, the track record of a state (or institution) in addressing a problem, its reliability in terms of motivation, and its capacity to revise the policy in case of failure.

Addressing the structural causes of the wrongs that EA wants to tackle, and targeting states and institutions too, means that EA should also consider the further issues mentioned in the previous section: what is the 'ideal of EA'? Are the duties we have duties of a non-ideal situation that would disappear in a just condition? How is the 'ideal of EA' accessible and stable? Should we only try to convince people to adopt EA's principles, or would it be desirable to coerce people to abide by these principles? Is the theory of effectiveness for public and institutional actors the same as the one EA uses for individual and private organizations? Answering all these questions – and many others – requires further theorizing and a broader dialogue between EA's typical focus on individual moral choices and the perspective of political philosophy. Although this might seem too long a detour for EAs, who are rightly concerned with the urgency of many practical issues, it should not be considered a waste of time for a movement that is committed to theoretically grounded social engagement. EAs would object that answering all these questions *is* a waste of time given that more theorizing along the lines I have suggested does not seem conducive to doing more good. In reply, this is not necessarily true to the extent that clarifying one's ideal serves the purpose of establishing the direction of the movement. Establishing this, as well as the other aspects mentioned above, is also necessary if the movement gains further support. To the extent that the movement expands its activities and influence, a better understanding of the compatibility of diverse goals and their influence on other societal challenges is a practical and theoretical necessity. In sum, the possible scaling-up of the movement also requires some theoretical choices on its foundation and compatibility with other social issues.

To conclude, regarding the question of whether EA can support structural changes, we may admit that in theory EA could justify the initiative to support changes in societal, political and structural issues. However, whether and how such changes may be endorsed by the current EA framework is uncertain because it all depends on the estimation of the probabilities of success, *plus* the weighing of the importance of possible outcomes. That very important changes may be justified despite their low probabilities of success is possible but unlikely within EA's framework, given that the (understandable) focus on effectiveness leads to a preference for goals that are more easily and clearly achievable. An expansion or reformulation of EA's conceptual tools and its idea of effectiveness along the lines suggested above may be a way to address the ambiguous attitude towards structural changes.[26]

[26] I am grateful to Chiara Cordelli, Rainer Ebert, Josh Milburn, Selina O'Doherty, Gianfranco Pellegrino and to two journal's anonymous reviewers for their helpful comments on previous versions of this article.

Cite this article: Zuolo F (2020). Beyond Moral Efficiency: Effective Altruism and Theorizing about Effectiveness. *Utilitas* **32**, 19–32. https://doi.org/10.1017/S0953820819000281