Centro
di Ricerca e
Documentazione
Luigi Einaudi

Federico Zuolo

EFFICIENTLY ETHICAL ON THE EFFECTIVENESS CONDITION IN EFFECTIVE ALTRUISM

workingpaper
N.7/2018

Gli Annali 2018

A cura di Matteo Jessoula, Beatrice Magni,
Nicola Riva, Maurizio Ferrera

ISBN 978-88-94960-00-6

Laboratorio di Politica Comparata
e Filosofia Pubblica

FEDERICO ZUOLO

EFFICIENTLY ETHICAL
ON THE EFFECTIVENESS CONDITION IN EFFECTIVE ALTRUISM

Centro Einaudi • Laboratorio di Politica Comparata e Filosofia Pubblica
With the support of Compagnia di San Paolo

Working Paper-LPF n. 7 • 2018

ISBN 978-88-94960-00-6 • Annali 2018

Federico Zuolo is Assistant Professor at the University of Genova, Dept. of Classics, Philosophy and History. He previously was post-doc fellow at the University of Trento, Pavia, and IUSS. He also was Alexander von Humboldt Foundation Senior Research Fellow at the University of Berlin (Freie) and Hamburg. His interests concern animal ethics, respect, equality, and the feasibility of normative political theories. He published an introduction to animal ethics (*Etica e animali. Come è giusto trattarli e perché*, il Mulino 2018), a new edition of Xenophon's *Hiero* (Carocci, 2012), a monograph on Plato (Academia Verlag, 2009), and a number of articles in such international journals as *Utilitas*, *European Journal of Political Theory*, *Journal of Applied Philosophy*, *International Journal of Philosophical Studies*, *Journal of Value Inquiry*, *Journal of Social Philosophy*, *Ethical Theory and Moral Practice*, *Social Theory and Practice*, *Philosophical Papers*, *Revue Philosophique de la France et de l'Etrager*, *The Ethics Forum*, *Historical Social Research*.

The **Comparative Politics and Public Philosophy Lab (LPF)** at Centro Einaudi is directed by Maurizio Ferrera and funded by Compagnia di San Paolo. It includes the **Welfare Laboratory (WeL)** and the **Bioethics Lab (La.B)**. LPF analyses the transformation of the political sphere in contemporary democracies with a focus on the relationships between policy choices and the value frameworks within which such choices are, or ought to be, carried out. The reference here is to the "reasonable pluralism" singled out by John Rawls as an essential feature of political liberalism.

The underlying idea is that implementing forms of "civilized" politics is desirable as well as feasible. And, as far as the Italian political system is concerned, it is also urgently needed, since the system appears to be poorly prepared to deal with the challenges emerging in many policy areas: from welfare state reform to the governance of immigration, from the selection criteria in education and in public administration to the regulation of ethically sensitive issues.

In order to achieve this end, LPF adopts both a descriptive-explanatory approach and a normative one, aiming at a fruitful and meaningful combination of the two perspectives. Wishing to foster an informed public debate, it promotes theoretical research, empirical case studies, policy analyses and policy proposals.

LPF • Centro Einaudi
Via Ponza 4 • 10121 Torino
Phone +39 011 5591611
segreteria@centroeinaudi.it • www.centroeinaudi.it

Efficiently Ethical.
On the Effectiveness Condition in Effective Altruism

Table of Contents

Keywords

Effective Altruism, Efficiency, Feasibility, Realism

ABSTRACT

EFFICIENTLY ETHICAL
ON THE EFFECTIVENESS CONDITION IN EFFECTIVE ALTRUISM

Effective altruism is a social movement that aims to improve the way we help others. By combining reflections on the ethics of giving with a more updated understanding and data of what is an effective choice, its proponents seek to ameliorate the impact that each person can have on the welfare of others, in particular the worst-off. It has been criticized for focusing on the consequences of disadvantage, thus overlooking its deep causes. In this paper 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.

EFFICIENTLY ETHICAL

ON THE EFFECTIVENESS CONDITION IN EFFECTIVE ALTRUISM

FEDERICO ZUOLO

1. INTRODUCTION: EFFECTIVE ALTRUISM, THEORY OF A PRACTICE AND PRACTICE OF A THEORY

Effective altruism (henceforth EA) is one of those few cases 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, that of doing the 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.

In virtue of its strong link between abstract moral principles and reality, EA must be assessed in a multifaceted manner. Many have discussed the moral premises of EA or its actual conduciveness to good outcomes. In this paper, instead, I provide a critical analysis of an underexplored part in EA, namely its effectiveness dimension. I will thus not address other fundamental issues regarding the ethics of giving, as to how much and why we ought to donate, to whom should we donate first, whether there is a conflict between the duty of impartiality and legitimate partial commitments (MacAskill 2016, McMahan (mimeo), Pummer 2016, Singer 2015), how effective altruists can get rid of the charge of falling into a "repugnant conclusion" (Pellegrino 2017) and so on. Rather, I will analyze the framework and presuppositions behind the effectiveness condition. Given the importance of effectiveness in EA it may be worth investigating this problem from a theoretical perspective, with a view to checking whether EA lives up to the theoretical ambitions that effective altruists set forth for EA. Hence, I will not level the classical, and frequently misused critique of non-applicability, namely that EA cannot realize the ideal in practice, an argument that by definition may be pressed against almost any moral perspective. Rather, I will try to understand 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 paper proceeds as follows. In the next section I provide a definition of effectiveness and try to better understand EA's conception of effectiveness. In the third section, I distinguish effectiveness and efficiency and bring light to EAs' penchant for efficiency. Then, I briefly mention the main critiques leveled against EA. In the fifth section, I bring to the fore 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 previously mentioned. The final section concludes and provides a sort of agenda for effective altruists to make EA more in line with a conceptually and practically deeper understanding of the commitments to effectiveness.

2. A Definition of Effectiveness

Given its theoretical formulation and ambitions, EA must be assessed *qua* theory, even when we discuss how it envisages practical outcomes. To do this we must formulate and analyze EA's account of effectiveness. In what follows, I provide a general definition of effectiveness of a theory and the correlate actions. Here, by theory I broadly understand a set of descriptive, methodological, evaluative and prescriptive statements that aim to interpret the world, assess it and guide individual and institutional actions. As effectiveness pertains to actions first, and only indirectly and figuratively to a theory, let's start with the effectiveness of actions.

> An action (A) is effective (E) with respect to a value (V) iif the state of affairs (SoA) brought about by (A) embeds (V).
>
> A(E,V) iff $(A \rightarrow SoA) \,\&\, (V \subset SoA)$[1]

This definition is binary because it is made ex post given the certainty of knowing what has in reality been effective. But in many cases we also need to estimate in advance the possible effectiveness of an action, hence we need to go probabilistic.

> 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).

[1] In what follows → is to be understood as indicating a causal connection, ⊂ should be understood as the standard operator of inclusion. Similarly, effectiveness for a theory can be formalized as follows:
$E(Th,V) = (A_{Th} \rightarrow SoA) \,\&\, (V \subset SoA)$.
Needless to say, these formalizations should be taken with a pinch of salt.

$$EE(A,V) = P(SoA \mid A) \ \& \ (V \subset SoA)^2$$

Here we have a scalar measure of effectiveness through its probability. This value may be used to compare the expected effectiveness of diverse theories or proposals.[3]

A couple of clarifications before proceeding. For the idea of effectiveness to make sense, first, (A) should be an intentional action that is causally conducive to (SoA). Otherwise, we have a state of affairs that is the unintended result of an action (or set thereof) that is mingled with other actions conducive to a state of affairs without determining it. If so, why should we be concerned with the effectiveness of a theory and in particular of the actions drawing on it if such actions were irrelevant to the determination of the (SoA) at stake?

Second, one might be worried that by saying "(V ⊂ SoA)" we actually commit a metaethical mistake in representing a set of values (V) included in a set of facts (SoA). According to some metaethical outlooks this is at least dubious and needs some explanation. In reply, by (V ⊂ SoA) I mean that (SoA) includes some (factual) features realizing a certain (V). How such a realization is metaethically possible does not concern us here. What these features are is variable. It may be a set of procedures (for instance, the value of equal treatment before the law is realized by legal provisions, plus the independence of the judiciary, plus the institution of legal aid, and so on), or a monetary provision (for instance, the value of freedom from exploitation might be realized by unconditional basic income).

Third, (A) may be may be congruent with (V) – if so we have a type of moralized effectiveness – or in contrast with (V) – if so we have a Machiavellian effectiveness. This distinction may be irrelevant in case an effective altruist is a thorough utilitarian, thus disregarding the value of integrity and considering involvement with bad actions morally justified if this is conducive to a good outcome. But it might be relevant for approaches to EA that are sensitive to the value of integrity.

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 resources? It seems rational to suppose that given our limited resources and abilities, the complexity of the world and the urgency of some problems, we'd better focus on problems we can expect that can be solved by agents about whom

---

[2] The same can be said of the expected effectiveness of a theory regarding the value and action it recommends: $E(Th,V) = P(SoA \mid A_{Th}) \ \& \ (V \subset SoA)$

[3] Building on this, we can assess the diverse probabilities of diverse theories and establish the best one. If we suppose that $P_{th1} < P_{th2} > P_{th3}$, then it follows that in terms of expected effectiveness we have to prefer Th2 in virtue of its higher probability.

we have reliable information. Four features seem obvious candidate to make sense of this.

The first two features are *directness* and *specificity of goals*. I define a goal as a (V ⊂ SoA) which a theory recommends as desirable. EAs insist 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 with the realization of vague goals for whose implementation we have no clear route? However, the way in which this reasonable general recommendation is actually understood by theorists of EA prescribes a rather restricted set of actions, namely preferably, albeit not only, the donation of as much as we can to some selected NGOs and charities.

The third and fourth features concern the measurability and the time frame of effectiveness. On EA's view, whether an action is effective should be easily *measured* and should be *short-term* traceable. The two features are linked: long-term goals are hard to assess. Measurability also is related to the theory of value that most of EAs assume (some form of welfarism) but here is to be understood as a matter of what is to be counted as a real achievement, which could make a state of affairs an effective altruistic result.

From this it follows that specific, short-term, measureable goals for whose implementation a direct action is available are to be preferred over general, long-term, complex goals whose implementation depends on many intermediate and indirect steps. Why is this so? It is so because the latter kinds of goals have an uncertain epistemic status in that it is difficult to establish how likely they are and what we ought to do to reach them.

These four 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 complicate. This is so for many reasons. 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 behaving), 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 people to 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 (not) to deal with the objective complexity of the world because it simply recommends to donate 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 regarding the type of goals to be pursued:

> Not any goal ($V \subset SoA$) 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 state of affairs (i) maximising the amount of value included: ($V_{MAX} \subset SoA_i$)
>
> In theory, we can devise only an action (or a restricted set thereof) that brings about such a state of affairs maximising the desired value: $\exists!A_i \rightarrow (V_{MAX} \subset SoA_i)$

3. 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, infact, 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 as follows. Effectiveness concerns the overall capacity of an action to actually 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 iff, given a fixed amount of resources (R), it maximizes the amount of (V) in (SoA) with respect to other feasible alternatives.

Or in formal terms:

> Among a set of possible actions ($A_{1\text{-}n}$), A is efficient iff A_i : Max ($|R| \rightarrow V_i$) & ($V_i \subset SoA$)

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 among competing scores of efficiency.

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 of effective altruists, and in particular Pummer and MacAskill, argue for the conditional duty to be efficient upon deciding to donate, that is using the resources to generate the maximum expected amount of output. MacAskill repeatedly mentions the criterion of cost-effectiveness, which is efficiency in other terms.[4]

Second, they seem 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 each monetary input received.

Third, they (in particular MacAskill 2016) recommend of us to assess the value of our actions – regarding donation but also the choice of our career – in marginal terms given the law of diminishing marginal returns, namely in terms 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 career of others.

Pallotta (2015) has argued against the idea that EA is about efficiency. But he takes too narrow an understanding of what efficiency means. By efficiency he means merely the organization or procedure which 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 efficiency includes the notion of goal, ($V \subset SoA$), which is what counts in practical matters.

If this interpretation is convincing, it shouldn't be seen as a disappointing fact. After all, as seen, efficiency is a form of effectiveness. But it is not the same given its specific focus on the maximization of a given amount of resources.

Now the question is whether this understanding of effectiveness lives up to its ambitions. Critics point out that one cannot really aim at improving the condition

[4] "We want to estimate what a charity achieves with a given amount of money, so our focus should always be on *cost-effectiveness* rather than just *effectiveness*." (MacAskill 2016, 135).

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. (Herzog 2016, Srinivasan 2015). By refusing to address these issues effective altruists fail to tackle the real underlying causes of the most pressing problems that effective altruists are presumptively committed to solve. 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 to the state service whose quality is likely to be diminished because of the lower pressure that people put on state agencies (Clough 2015). 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 (Wenar 2010).[5] In sum, such critics claim that EA lacks insight in the structural causes of moral wrongs because EA is concerned only with the effects of deep problems. Hence, by overlooking the systemic and long-term issues EA does not live up to its practical commitments.

Hence, private donations won't do. These remarks seem in many cases plausible. But to be fair towards EA, EAs also have a point in focusing only on actions that can actually bring about real and measurable, despite minor, improvements. In one sense EAs' approach is likely to be very effective, at least if we compare it to other available alternatives. Indeed, effective altruists seem to rely on the assumption that we should practically care only about issues for which we can have a sizeable and expectedly direct impact. To be true, 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 regarding such complicated issues?

EAs seem hyper-realists while also being very moralist. On the one hand, they are in a sense hyper-realists and somewhat pessimists 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 out of our reasonable and foreseeable control. But, on the other hand, in so doing they recommend to do as much as we can, thus being very demanding moralists.

This seeming divergent attitude has a ground. We have seen that effective altruists are committed to a sort of efficient management of the (scarce) resources that have to be devoted to donation in order to yield as much good as possible. This includes the commitment to efficiency when it comes to the management of personal

[5] Here I only focus on critiques to EA's capacity to recommend really effective actions. Of course, there are a number of more principled critiques. On these see, McMahan (unpublished).

concern, as if one had a limited amount of resources of moral concern at her disposal.

However, this focus on efficiency is not necessarily the best way to pursue effectiveness in general. After all, there are many relevant causes whose advancement requires and has required actions that we would hardly qualify as efficient. If effectiveness is usually understood in the light of efficiency, how EA's theory of effectiveness could cope with the hardly controversial statement that some of the most important social changes that put an end to severe moral wrongs were brought about by social activism that apparently did not follow EA's prescriptions? For instance, Gandhi's pacifist movement for India's independence, M.L. King's march for the rights of Afro-Americans, and Mandela's fight for the end of apartheid regime were all complex 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 a faith was both a hope in the final success and a staunch commitment to the morality of their cause, whatever the probability of success could be. 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 exaggerate 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? 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 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? Isn't EA's way of measuring expected effectiveness not only too timid but also too conservative when it comes to complex social changes?

To be sure, MacAskill (2016, 114-121) claims that EA is not reducible to charity donations because EA can also justify and recommend actions that are not easily quantifiable if the expected gains are high (like in the case of political careers). However, in making reference to a political career and other non-easily quantifiable expected goals, MacAskill seems more to be paying a lip service to this hypothesis than truly recommending such options according to the EA's framework.

Moreover, MacAskill only considers the individual perspective of career and donation, not the overall systemic question. As Gabriel (2016, 13) convincingly points out effective altruism is wedded to the logic of changing frequently the priorities and addressees of donations in virtue of the varying marginal effectiveness. Hence, it seems ill suited to the pursuit of goals requiring longstanding, stubborn and patient activism, which was eventually effective. In sum, given that big social changes did not seem to follow EA's theory of effectiveness as efficiency, shouldn't we revise it?

If, instead, effective altruists would prefer sticking to their theory in which efficiency, measurability and directness play a paramount role, they should rather look for an alternative ground to back it, i.e. a justification for this account which does not rest only on its effectiveness-value. Paradoxically as it may seem a Stoic interpretation of EA's focus on 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. What depends on us is 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.[6] From this it follows that our control should be directed only towards our inner dispositions, not towards what happens in the world, which cannot be controlled by us. Of course, Stoic's concern with one's inner sphere does not mean retreat 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 to be primarily focused on those things that you can control directly. Of course, unlike the Stoic 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 her 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 to achieve it at best, is not under one's control and hence should be, so to say, "externalized" to meta-charities and NGOs.

[6] "Some things are up to us and some are not up to us. Our opinions are up to us, and our impulses, desires, aversions—in short, whatever is our own doing. Our bodies are not up to us, nor are our possessions, our reputations, or our public offices, or, that is, whatever is not our own doing". Epictetus 1983, §1, p. 11.

This Stoic interpretation is suggestive but I don't want to overstate it. To conclude, whatever the best interpretation of effectiveness in EA, EA has a peculiar strategy to deal 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.

4. Effectiveness and Other Cognate Notions

In sum, EA's account of effectiveness seems too reductive in seeing effectiveness merely as a matter of efficiency, and in not being equipped to drive structural social changes. But, is it necessarily so? Can we broaden EA's account of effectiveness so as to address these two important challenges? To pursue this goal, I claim that 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 distinctions and arguments laid out in the debate on 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 rough 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. It has been asked what appropriate level of fact-sensitivity a theory of justice should have. Following, G.A. Cohen's critique of Rawls, the issue of sensitivity to facts has been understood as to whether and how much, in order to be valid and sound, principles of justice should depend on (social or natural) facts. Answers have differed, ranging from those holding that first principles of justice should be independent of facts (Cohen 2008), to those arguing that facts are not only relevant 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 (Rossi 2016).

Related but not identical to this question, there has been the distinction between ideal and nonideal 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 favorable conditions, the parties comply with the principles of the theory. This assumption is 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 (Rawls 1971). Accordingly, one may ask what the duties are, given a situation of non-compliance (non-ideality) of the parties. For instance, are the people required to abide by the same political obligation if they find themselves in a situation where other parties are not compliant or the institutions are far from the well-ordered society?-The other sense of the distinction between ideal and non-ideal theory is more in line with the issue of fact-sensitivity we have just presented. 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 condition of the least fact-sensitivity, where highly abstract principles are presented and defended, to the condition of a theory that includes and mirrors a number of empirical and social facts (Valentini 2016). Hence, the question is about which level of fact-sensitivity is appropriate to discharge the functions of a normative political theory in terms of its capacity to both guide actions and be justified independently of the varying contexts.

Second, effectiveness is related to feasibility. By definition what is effective also is feasible. Feasibility is a modal concept that concerns whether a state of affairs can be brought about by individual or collective actions. As it has been conceptualized by Lawford-Smith and Gilabert (2012), (political) feasibility can be both understood as a binary concept, and a scalar concept. 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 that an action has in order to bring about the desired state of affairs given some soft-constraints (cultural, economic, social). The types of feasibility impact differently on our duties and practical recommendations. 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 alternative. Measures of feasibility are relevant to the ex ante estimation of whether and how much an action might be effective. Indeed, the ex post assessment of an NGO's action is not always available or telling, in particular if EAs are interested in going beyond short-term and specific-scope kinds of actions.

Third, a further distinction that is relevant here is that between the access-dimension of feasibility and the stability-dimension of feasibility (Lawford-Smith 2013). 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 at 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 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 relevant to the relation between facts and values, ideals and reality and our practical commitments towards a just world is that of reconciliation. Rawls (2001, p. 3) claimed that one of the tasks of a normative theory is also that of reconciliating ourselves with social reality, by showing that current rules and institutions, and their history display a rational and justifiable form. The importance of reconciling is that of showing that what we seek through normative theory is at least partially present in reality, thus pointing out that reality is not completely immoral or irrational, and that our endeavors to realize justice more fully are not doomed or inane.

5. 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 (badness of suffering, poverty etc.) to be self-evident, unquestionable and accessible to all well-meaning people. In being universal and obvious they are independent of circumstances. However, the practical and theoretical approach of EA's theory is set up in a markedly non-ideal fashion. The moral principles that EA recommends are first-order principles that are thought to be valid and compelling in any condition, be it ideal or fully non-ideal. Yet, they are all principles regarding individual morality because all EAs are implicitly or explicitly committed to some form of consequentialism which is scarcely institutional, namely a view where institutions are merely instruments to deliver the only valuable good (people's well-being). However, whether and how much moral principles are binding and action-guiding also depends on the favorability 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 and how much people are in need and the extent to which other people are also complying with this principle. A world of altruistic donators would not only be impossible, qua too demanding, but also qua practically

ineffective insofar 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. (More on this below).

Regarding feasibility, EA mostly adopts an ex post measure of success. Effective altruists typically recommend relying on the effectiveness-assessment of meta-charities that evaluate the reliability of NGOs and charities. However, ex post provide a biased assessment of feasibility. 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 the domain of what we already know and what we have already done.

Regarding the binary and scalar understanding of feasibility, in EA we find no clear distinction between hard and soft-constraints because no clear picture of the overall ideal of society is put forward. Perhaps, it is more appropriate to say that implicitly EAs deal with soft-constraints because investigating what the hard constraints are is a theoretical enterprise which is left out of the picture.

In a sense, EA's conception of E is surprising and paradoxical. On the one hand, it is rather "moralistic" in that it targets individuals' set of motivations and attitudes requiring that each devotes a significant portion of one's income to donation. And it may condemn those who fail to do so: while Pummer (2016) formulates a conditional duty to donate, Unger (1995), extending Peter Singer's old argument of the pond, defends a general duty to donate on the part of the well-off. On the other hand, it may seem extremely conservative because it does not tackle institutional issues regarding power and the mechanisms that create current enormous inequalities and differences of opportunities etc. It is not clear whether this lack of concern for structural issues depends on a skeptical expectation that such issues won't be changed any soon, *plus* an overall commitment to be effective in a near sizeable future, or whether this lack of concern hides an overall appreciation of current institutional structures. Or whether this depends on a methodological stricture that impedes the understanding of issues that do not pertain to individuals' moral concern insofar as they go beyond individuals' capacities.

In sum, this overall conception of effectiveness is demanding but non-ideal, focusing on the most pressing problems without being fully concerned with the causes of these problems. Is it neither a radical view, nor a conservative one. In a slogan, it's a sort of *remedial effectiveness*, in that it is only concerned with the salient expressions of problems without addressing the underlying structures thereof. (EAs, I'm sure, would not take it as a derogatory comment).

Building on this we may try and outline what EA could say on these other issues related to effectiveness even though they have not been outlined yet. Given the lack of clarity on EA's overall commitment towards the goodness of current states of affairs and the need to go beyond it, such a set of questions does not seem out of place.

First one may ask: What EA could say about access-feasibility? And access to what? 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 state of affairs, we should ask what this better state of affairs consists in. On this we have no clear clue, as effective altruists 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 as to why and how these situations are moral problems. So, 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 what these problems are, what are the priorities and how a society without these problems would look? This is hardly so, and effective altruists would be better having at least a working idea of what kind of end-state they are working for.

If we suppose that effective altruists 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? What kind of changes would be required depends on other issues. People like Singer would not accept a change of capitalism's basic system (see below). (In passing it is worth remarking that Singer despite having written and analysed a large number of issues has always been scarcely concerned with institutional issues in the proper sense.)–Some other activists and commentators have held that EA actually "loves systemic change"[7]. However, it is not clear what it means, whether a more democratic society, a more affluent one, or a vegan one. 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, such a specification of the most pressing targets is in fact missing from the picture.

Let us turn next to the stability question. On the one hand, this question depends on whether EA demands a transformation of our current world which, as we have just seen, we don't know in detail. 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

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

world would look like if all, or at least a conspicuous number of people would follow EA's predicaments, even within the current institutional system. Schmidtz (2000) has argued that, if people were to strictly follow the ethics of giving – which is for our purpose here the same as EA –, the current world would actually collapse because people would stop spending money for a vast array of issues that would be considered futile compared to the urgency of such issues as famine, fatal diseases etc. But if so, large portions of our economic system would simply be destroyed insofar as there would be no market. This is clearly undesirable from many perspectives, including consequentialism to which most of effective altruists subscribe. Is EA a universal account that applies to any realistic domain or is it applicable only under very non-ideal conditions? What is EA's functional equivalent of Rawls's circumstances of justice?

Moreover, what if we could convince most people to become effective altruists? What would a world embedding the principles of EA look life? To make it more functioning than in Schmidtz's imaginary world, we ought also to consider some institutional issues. 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 where all or most of 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 amount to an ethical ideal with still little institutional framework. To tease out its features, it is perhaps helpful to compare it with another quite moralistic and demanding theory: G.A. Cohen's idea of 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). (Tomlin 2010)

In Cohen's just society, morality demands that one forsake market driven (monetary) incentives to be mostly productive and contribute to the social cooperation. In this society people should accept to be paid less than their real productiveness and valuable contribution to social cooperation, in virtue of people's commitment to the egalitarian ethos. Effective altruists take the opposite route and claim that market driven incentives (and institutions) should be employed to maximize social production, so that bigger parts of it can be devoted to donations to the most needy. It is unclear, though, whether such a motivation can remain purely a matter of individual ethical behaviour in a framework where the

institutions not only do not reinforce it but also cause parts 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 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, standard roles and social functions, while giving new meaning to them and requiring an important change in the personal advantage that one takes from one's position. So unlike other theories aiming at reconciliation, in EA we do not know whether there is anything intrinsically good in how current states of affairs are organized, but some reconciliatory effect is at least implicit in EA's theories. Well, 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" (Singer 2015, p. 50). Hence, the worries about being complicit in a rotten system should be weakened because capitalism's overall consequences are positive.

6. Conclusion. As a Sort of Agenda

In conclusion, let me sketch out some considerations on what effective altruists could do to complete their account of effectiveness.

If what I've argued is correct, EAs interpret effectiveness in terms of efficiency of individual contributions. I don't claim this perspective is wrong, it just seems limited because (i) it cannot track structural social changes, (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. In sum, it is understandable if we only take the point of view of what individuals *qua* individuals ought to do. However, as a sort of constructive proposal effective altruists might take into account the possibility of adopting the perspective of (private or public) collective agents, and of recommending the engagement in activities whose effectiveness has not already been assessed by something like a meta-charity.

Practical urgency should not shy effective altruists away from trying to understand the deeper causes of the wrongs they want to address in practice. This entails understanding the institutional causes of famines, poverty and deprivation.

Practically addressing these problems is of course more risky and the possibility of success more feeble than other more direct and specific goals. How should the theory of effectiveness change?

It should perhaps not only focus on individuals' choices. This is to an extent inevitable and part of the capacity of EA to convince people by showing the possibility that each can have an impact. However, without dismissing the importance of individual and private initiatives, durable and fundamental change cannot be brought about without institutions and states. How can effective altruists estimate the probability of success with 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 that we have 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? Is it sufficient to rely on the convictions of people and on their individual initiatives or should we involve public actors? Is the theory of effectiveness for public and institutional actors the same as the one EA is employing for individual and private organizations? Answering all these questions requires further theorizing which, though, should not be considered as a waste of time for a movement that is committed to a theoretically grounded social engagement.

REFERENCES

Berkey, B. (2017). "The Institutional Critique of Effective Altruism." *Utilitas*, forthcoming.

Clough, E. (2015). "Effective Altruism's Political Blind Spot." *Boston Review* 14 July.

Cohen, G. A. (2008). *Rescuing Justice and Equality*. Cambridge (MA)/London: Harvard University Press.

Epictetus. 1983. *Handbook of Epictetus*. Trans. by N. White. Indianapolis/Cambridge (UK): Hackett.

Gabriel, I. (2016). "Effective Altruism and its Critics." *Journal of Applied Philosophy*, online first, doi: 10.1111/japp.12176.

Gilabert, P., Lawford-Smith, H. (2012). "Political Feasibility: A Conceptual Exploration." *Political Studies* 60: 809-825.

Herzog, L. (2016). "Can 'Effective Altruism' Really Change the World?" *OpenDemocracy*, 22 February. Available at: https://www.opendemocracy.net-/transformation/lisa-herzog/can-effective-altruism-really-change-world, last access 10 March 2017.

Lawford-Smith, H. (2013). "Understanding Political Feasibility." *The Journal of Political Philosophy* 21(3): 243-259.

MacAskill, W. (2016). *Doing Good Better. Effective Altruism and How You Can Make a Difference*. London: Guardian Books.

Pallotta, D. (2015). "'Efficiency' Measures Miss the Point." *The Effective Altruism Handbook*, a cura di R. Carey, 32-34. Oxford: Centre for Effective Altruism.

Pellegrino, G. (2017). "Effective Altruism and the Altruistic Repugnant Conclusion." *Essays in Philosophy* 18(1): 1-24.

Pummer, T. (2016). "Whether and Where to Give." *Philosophy and Public Affairs* 44(1): 77-95.

Rawls, J. (1971). *A Theory of Justice*. Cambridge (MA): The Belknap Press of Harvard University Press.

—— (2001). *Justice as Fairness. A Restatement*. Cambridge (MA): The Belknap Press of Harvard University Press.

Rossi, E. (2016). "Facts, Principles and (Real) Politics." *Ethical Theory and Moral Practice* 19: 505-520.

Schmidtz, D. (2000). "Islands in a Sea of Obligation: Limits of the Duty to Rescue." *Law and Philosophy* 19(6): 683-705.

Singer, P. (2015). *The Most Good You Can Do. How Effective Altruism is Changing Ideas About Living Ethically*. New Haven/London: Yale University Press.

Srinivasan, A. (2015), "Stop the Robot Apocalypse." *London Review of Books* 37(18): 3-6.

Tomlin, P. (2010). "Survey Article: Internal Doubts about Cohen's Rescue of Justice." *The Journal of Political Philosophy* 18(2): 228-247.

Unger, P. (1996). *Living High and Letting Die. Our Illusion of Innocence*. Oxford/New York: Oxford University Press.

Valentini, L. (2016). "On the Messy "Utopophobia vs Factophobia" Controversy: A Systematization and Assessment." *Political Utopias: Contemporary Debates*, a cura di M. Weber, K. Vallier, 11-33. Oxford/New York: Oxford University Press.

Wenar, L. (2010). "Poverty is No Pond. Challenges for the Affluent." *Giving Well: The Ethics of Philantrophy*, a cura di P. Illingworth, T. Pogge, L. Wenar, 104-32. Oxford: Oxford University Press.