

Conditional obligations, moral theory, and the inference to effective altruism

A Reply to Pummer and Horton

Thomas Sinclair

1. A new argument for effective altruism

Effective altruists believe in “doing the most good with whatever resources you have,” and advocate “the use of high-quality evidence and careful reasoning to work out how to help others as much as possible” in order to achieve that goal.¹ Since these claims may seem very demanding, some philosophers suspect effective altruists of an underlying commitment to utilitarianism, which would provide an obvious foundation for such demandingness.²

For many people, however, effective altruism is chiefly understood as the less radical proposal that we should direct our charitable donations to the charities identified by such organizations such as GiveWell as most cost-

¹ William MacAskill, *Doing Good Better* (London: Guardian Books, 2015), p. 15; The Centre for Effective Altruism, ‘Introduction to Effective Altruism’, URL=<<http://www.effectivealtruism.org/articles/introduction-to-effective-altruism/>> (accessed 15 March 2017).

² See e.g. Amia Srinivasan, ‘Stop the Robot Apocalypse’, *London Review of Books* vol. 37, no. 18 (24 September 2015); Iason Gabriel, ‘Effective Altruism and its Critics’, *Journal of Applied Philosophy* vol. 34, no. 4 (2017).

effective rather than to others.³ This proposal is compatible with a range of views about the amount or proportion of our resources that we are morally required to give away. What's distinctive about effective altruism, on this less radical interpretation, is the claim that *if* we make any charitable donations at all, we are morally required to do as much good as possible with them. Call this the *conditional obligation of effective altruism*.

This less radical interpretation of effective altruism has some practical and theoretical attractions for proponents of efficient charitable giving. It appears to avoid standard 'demandingness' objections to utilitarianism, since it does not assume that charitable giving is morally required to begin with.

Nevertheless, it casts efficiency in charitable donation as a moral requirement. And yet it allows effective altruists to sidestep the difficulty of identifying a non-arbitrary quantity or proportion of one's resources that one is morally required to devote to others. Perhaps there is some such quantity or proportion. But if there is also an independent obligation of efficiency conditional on optional donation, then effective altruists can argue cogently that we ought to be doing far more good even if the level of giving that is unconditionally required cannot be determined.

Even on this less radical interpretation, however, the truth of effective altruism cannot be assumed. Most people do not think they are doing something morally wrong when they make donations to charities not in GiveWell's top 5. So effective altruists will have welcomed an argument for the

³ GiveWell, 'Our Criteria for Top Charities',
URL=<<http://www.givewell.org/how-we-work/criteria>> (accessed 4 April 2017).

conditional obligation of effective altruism that has recently been made in this journal by Theron Pummer.⁴

Pummer defends the conditional obligation of effective altruism on the basis of intuitively correct judgments about ‘rescue cases’. These are situations in which an agent can save either a smaller or a larger number of people from harm with no difference in cost to herself, but in which that cost is in any case so great as to make it permissible for her not to save anyone. Pummer’s defense instantiates the following general form of argument, which I’ll call the *Rescue Argument for Effective Altruism* (or Rescue Argument for short).

1. As our intuitive judgments about rescue cases show, there are conditional obligations to rescue as many people as possible in some cases if one is going to rescue any people at all.
2. We can infer general principles of conditional obligation from an analysis of the moral mechanism that explains our judgments about the rescue cases.⁵

⁴ Theron Pummer, ‘Whether and Where to Give’, *Philosophy & Public Affairs* vol. 44, no. 1 (2016), pp. 77–95.

⁵ Arguments resembling steps 1–2 have also been made by Derek Parfit, ‘Future Generations: Further Problems’, *Philosophy & Public Affairs* vol. 11, no. 2 (1982), p. 131; Shelly Kagan, *The Limits of Morality* (Oxford: Oxford University Press, 1989), p. 16; Ulla Wessels, ‘Beyond the Call of Duty: The Structure of a Moral Region’, *Royal Institute of Philosophy Supplement* 77 (2015), pp. 87–95; Dale Dorsey, *The Limits of Moral Authority* (Oxford: Oxford University Press, 2016), pp. 110–1; and Tina Rulli, ‘Conditional Obligations’, unpublished manuscript. These authors don’t go on to argue for the inference to a conditional obligation

3. The general principles imply a conditional obligation of effective altruism.

My aim in this article is to rebut Pummer's version of the Rescue Argument, alongside a different version advanced by Joe Horton.⁶ I set out Pummer's and Horton's arguments in more detail in section 3. As I explain in sections 4–5, both Pummer's and Horton's explanations of the intuitively correct judgments about rescue cases are strictly speaking non-consequentialist, because they concede the existence of options not to maximize the good. But they nevertheless presuppose a controversial conception of the moral domain and moral theorizing that is characteristic of consequentialism; as I put it, their non-consequentialism is only half-hearted. In sections 6–7, I explain how a more thoroughgoing non-consequentialist can account for the intuitively correct judgments about rescue cases. Then I give reasons to prefer the thoroughgoing approach to its half-hearted counterpart. I conclude that we should reject Pummer's and Horton's defenses of effective altruism. Finally, in section 8, I discuss the possibility of a thoroughgoing non-consequentialist version of the Rescue Argument. I suggest that constructing such an argument is more challenging, and that its conclusions will be more limited, than effective altruists might hope.

of effective altruism in particular, but in some cases they come very close to doing so. My arguments in what follows will also apply to much of what they say.

⁶ Joe Horton, 'The All or Nothing Problem', *Journal of Philosophy* vol. 114, no. 2 (2017), pp. 94–104.

2. Conditional obligations

I begin with a more detailed account of conditional obligations. Conditional obligations are moral requirements that are violated by failure to perform the required act only when that failure occurs under particular conditions.

Consider, for example, the moral obligation of gratitude. This is a conditional obligation, violated by a failure to be grateful only under the condition that someone has intentionally benefited you. By contrast, the obligation to treat others as ends in themselves is unconditional, and thus is violated by any failure to treat others as ends in themselves, no matter what the circumstances.

Though they may sometimes be stated in similar terms, conditional obligations are not hypothetical imperatives in Kant's sense. Hypothetical imperatives specify rational requirements that are conditional on some purpose that the agent has. For instance, if I have the aim of winning the lottery, the imperative "Buy a lottery ticket" applies to me. It would cease to apply if I abandoned the aim of winning the lottery, because its application is conditional on my having that aim. Conditional obligations differ from hypothetical imperatives in two ways. First, conditional obligations are not merely rational but moral requirements, failure to comply with which typically makes moral 'reactive attitudes' such as blame, resentment, and indignation appropriate. Hypothetical imperatives are not in general like this. If I don't buy a lottery ticket despite having the overriding aim of winning the lottery, you may chide me for my irrationality but you cannot appropriately *blame* me. Second,

whereas hypothetical imperatives always make the imperative conditional upon some purpose the agent has, conditional obligations need not be conditional upon such a purpose. It's not in general the case that one can evade the requirement specified by a conditional obligation by abandoning some purpose or aim that one has, since the condition in a conditional obligation is not generally that the agent has some particular purpose or aim.

3. Two Rescue Arguments

A. *Pummer*

Pummer's version of the Rescue Argument begins with the following example:

Arm Donor There is one innocent stranger stuck on track A, and a runaway train headed straight toward her. There are one hundred innocent strangers stuck on track B, and another runaway train headed straight toward them. If you do nothing, all one hundred and one people will soon die. However, you can use your arm as a train-stopper; those tied to the tracks are far enough away that putting your arm on either track would cause the relevant train to slow down and come to a complete stop before reaching anyone in its path. That is, if you place your arm on track A, you will stop the train on that track in time and save the one, and, if you place your arm on track B, you'll stop the train on that track in time and save the hundred. You have no other means of saving any of

these people. Since the two tracks are twenty feet apart, you will not be able to sacrifice more than one arm in time to save all one hundred and one. Assume that losing your arm is a large enough cost to you to make it not wrong not to incur this cost, whether that is in order to save one life or one hundred. All other things are equal.⁷

In *Arm Donor*, Pummer thinks, it is not wrong to do nothing – that is, not to sacrifice your arm. It is also not wrong to place your arm on track B, saving the hundred innocent strangers tied to that track. But it *is* wrong to place your arm on track A, since at no extra sacrifice you could have saved 100 rather than only one.⁸ This judgment about *Arm Donor* is, according to Pummer, best explained and justified by the following principle:

Avoid Gratuitous Worseness: it is wrong to perform an act that is *much* worse than another, if it is *no* costlier to you to perform the better act, and if all other things are equal.⁹

This is the second step in the Rescue Argument. The principle does not, however, condemn rescuing nobody in *Arm Donor*, because although rescuing nobody may be much worse in terms of lives saved, it is also much less costly to

⁷ Pummer, 'Whether and Where to Give', p. 83.

⁸ Pummer, 'Whether and Where to Give', p. 83. As Pummer notes, Derek Parfit and Shelly Kagan make similar claims about similar cases, as does Tina Rulli; see note 6 above.

⁹ Pummer, 'Whether and Where to Give', p. 84.

you. So *Avoid Gratuitous Worseness* is compatible with the existence of moral options, such as the option of doing nothing that you have in *Arm Donor*.

As Pummer observes, we can understand the combination of *Avoid Gratuitous Worseness* and the existence of moral options in terms of conditional obligations.¹⁰ The existence of moral options may exclude an obligation to help others, but if you *do* help them, you *must* perform the much more helpful of two acts if it is no more costly to do so, other things being equal. This is a general conditional obligation. And it is easy to see how one might infer from it a conditional obligation of effective altruism: the third step in the Rescue Argument. If there is this conditional obligation, and if you are going to make a charitable donation at all, and if some charities convert resources into benefits far more efficiently than others, then you must give your money to one of the far more efficient charities, other things being equal—even if you may permissibly make no donation at all.

Refinements of *Avoid Gratuitous Worseness* might yield a more stringent conditional obligation of effective altruism—for instance, by making it wrong to give to even a slightly less efficient charity, or requiring you to give to a more efficient charity even at slightly greater cost to yourself if you’re going to give at all.¹¹ But the essentials of the defense of effective altruism are already clear.

¹⁰ Pummer, ‘Whether and Where to Give’, p. 86.

¹¹ See Pummer, ‘Whether and Where to Give’, pp. 91–95. For this reason, Pummer includes the qualification ‘weak’ in his name for *Avoid Gratuitous Worseness*. Since it does not matter for my purposes whether, on Pummer’s account, the correct principle is this one or a stronger one, I omit the qualification.

B. Horton

Horton's argument also begins with an example:

Collapsing Building Two children are about to be crushed by a collapsing building. You have three options: do nothing, save one child by allowing your arms to be crushed, or save both children by allowing your arms to be crushed.¹²

As Horton observes, it seems intuitively that you are morally permitted to do nothing, given the sacrifice that saving the children would involve for you. Yet it would be wrong to save only one child, given that it would require no greater sacrifice to save both.¹³

As we saw, Pummer arrives at *Avoid Gratuitous Worseness* by generalizing from intuitive claims about what you may and may not do in *Arm Donor* and similar examples. By contrast, Horton proceeds by proposing a broadly contractualist approach that makes sense of intuitive claims about *Collapsing Building*, and then generalizing from that. He writes:

It is very plausible that if our acts are not justifiable to the people whom they affect, these acts are wrong. And plausibly, it matters not simply that our acts are justifiable to these people, but also that they are justifiable in a

¹² Horton, 'The All or Nothing Problem', p. 94.

¹³ Horton, 'The All or Nothing Problem', p. 94.

way to which we could reasonably appeal. If you were not willing to save either child, you could reasonably appeal to having to sacrifice your arms as a justification for not saving both. But [if] you are willing to save one child, you cannot reasonably appeal to this justification. Because there is no other adequate justification for not saving both children, there is no adequate justification to which you could reasonably appeal. So it follows, on this view, that [if] you are willing to save one child, you ought to save both.¹⁴

This analysis suggests the following general position. Even though you may be permitted not to confer some benefit on others because of the large sacrifice that doing so would involve, you must nevertheless confer this benefit rather than confer a smaller benefit, if you are willing to make a sacrifice to confer the smaller benefit and the sacrifices involved are of the same or similar size. This is because you cannot reasonably appeal to the sacrifice involved to justify not conferring the larger benefit, given that you are willing to make it to confer the smaller benefit.¹⁵

There is an exception: if you have “adequate agent-relative reasons” to favor the outcome in which you bring about the smaller benefit, then it is not wrong to bring about that smaller benefit.¹⁶ So, for instance, if you face a choice between saving your friend at the cost of losing your legs and saving two strangers at the same cost, then you may save your friend. But the

¹⁴ Horton, ‘The All or Nothing Problem’, p. 97.

¹⁵ Horton, ‘The All or Nothing Problem’, p. 98.

¹⁶ Horton, ‘The All or Nothing Problem’, p. 98.

contractualist explanation makes sense of this, since you can reasonably appeal to the fact that it is your friend in justifying your actions to the two strangers.

The contractualist explanation also appears to license an inference to the following general principle—step 2 in Horton’s version of the Rescue

Argument:

Optimific Altruism: if the only adequate justification for not bringing about an outcome O is that it requires a sacrifice S, and we are willing to make a sacrifice that is not significantly smaller than S to bring about an outcome that is not significantly better than O, and we do not have adequate agent-relative reasons to favor this other outcome, we ought to bring about O.¹⁷

Once again it is easy to see how one might go on to infer the conditional obligation in the final step, as Horton does. According to this conditional obligation, if we are willing to give to charity, unless we have adequate agent-relative reasons to give to a suboptimal charity, we ought to give to one of the charities that would use our gift to do the most good.¹⁸

¹⁷ Horton, ‘The All or Nothing Problem’, p. 99. I assume that for the requirement specified by *Optimific Altruism* to come into effect, the outcome that we are willing to make an S-sized sacrifice for must be significantly worse than O, not just ‘not significantly better’.

¹⁸ Horton, ‘The All or Nothing Problem’, pp. 102–4. *Optimific Altruism* differs from the conditional obligation defended by Pummer in that it makes the obligation of effective altruism conditional upon mere *willingness* to bear the

Refinements of *Optimific Altruism* might involve clarifying that other moral claims, such as those generated by prior promises, may be among the adequate agent-relative reasons to bring about a suboptimal outcome, as well as perhaps that non-agent-relative reasons such as considerations of culpability, voluntariness, or desert may count as adequate justifications for bringing about the suboptimal outcome. For instance, the greater deservingness of someone in need might in some circumstances justify providing aid to her rather than to another person who is in greater need but less deserving.¹⁹ But the essentials of the defense of effective altruism are once again clear.

4. A shared conception of the moral domain

Both Pummer's and Horton's versions of the Rescue Argument are officially non-consequentialist. Nevertheless, they share a conception of the moral domain that is characteristic of consequentialism.

relevant cost that makes it permissible not to act altruistically in the first place, rather than on the *choice* to bear that cost and act altruistically. Whether Horton's or Pummer's position here is more plausible depends partly on what, exactly, mere willingness without choice consists in and under what conditions its presence suffices to discount the reasonableness of appealing to the costs that one is willing to bear in a justification of failure to benefit another. But these details do not matter for my purposes, and I focus on the central cases covered by both Pummer's and Horton's positions, in which the willingness is manifested in a choice to benefit others.

¹⁹ (Acknowledgment removed.)

Thus Pummer argues that “there are *moral options*, according to which it is at least sometimes not wrong for you to act in a way that would result in less good [as judged from an impartial perspective] than if you had performed some other act instead.”²⁰ The recognition of moral options is sufficient for non-consequentialism. But the basis of moral options, on Pummer’s account, is simply that the cost to the agent of a consequentialist requirement to promote goodness over a certain threshold would be too great.²¹ So moral options are effectively conceived as a domain of licensed non-optimific action against a background presumption that action ought to be optimific. This is consequentialist in broad outlook, even though it allows for the presumption’s defeat in cases of excessive cost to the agent.

The same conception underpins Horton’s account, although it is less obvious. *Optimific Altruism* allows that the conditional obligation to bring about a better outcome may not apply if there are adequate agent-relative reasons to favor the worse outcome, and it may also allow for the possibility of other considerations that preclude its application, as we saw at the end of section 3. Furthermore, *Optimific Altruism* is compatible, strictly speaking, with the possibility that some failures to bring about the optimific outcome trigger no demand for justification at all: *Optimific Altruism* is addressed only to the case in which a justification for bringing about an optimific outcome is needed.

²⁰ Pummer, ‘Whether and Where to Give’, p. 79.

²¹ Pummer, ‘Whether and Where to Give’, p. 80. The same analysis is offered in more detail by Rulli. See Rulli, ‘Conditional Obligations’.

But Horton's own explanation of the exceptions and his later inference to the conditional obligation of effective altruism suggest that he takes failure to bring about the better outcome in general to trigger a demand for justification on pain of moral wrongdoing, a demand that only considerations of cost or adequate agent-relative reasons can meet.

Thus, to illustrate the exception, Horton gives an example in which saving your own child is bringing about the lesser benefit and saving two others is bringing about the greater.²² You have, he grants, adequate agent-relative reasons to save your own child. Throughout his article, almost all the adequate agent-relative reasons that Horton mentions—he also describes them as “personal reasons”—are considerations of agential sacrifice.

And if Horton thought a failure to bring about impartially good outcomes did not generally trigger a demand for justification, then further explanation of the triggering conditions for the conditional obligation of effective altruism would be needed, since Horton takes the conditional obligation to require doing impartial good rather than, say, fulfilling duties of humanitarian need or social justice. So it is reasonable to assume that he does take failure to bring about impartially good outcomes to trigger the demand. Indeed, this is what the contractualist framework should lead us to expect: contractualism is not normally thought to exclude any particular type or occasion of possible benefit to an individual as an input into contractualism's justificatory apparatus.

So Horton too seems to presuppose a conception of moral options according to which the only way one can justify non-optimific action is by invoking an

²² Horton, ‘The All or Nothing Problem’, pp. 98–9.

agent-relative prerogative grounded in considerations of something like demandingness. Again, moral options are effectively conceived as a domain of licensed non-optimific action against a background presumption that action ought to be optimific. Even if the range of reasons to which one may appeal for failing to bring about the greater benefit were widened so as to include duties to others or reasons of culpability, voluntariness, or desert in the way that I described at the end of section 3, the background presumption of an optimizing default would remain. The only thing added would be a kind of freedom to take account of *other* moral demands.

Given this shared conception, it's no surprise that when the exemption for personal cost does not apply because the agent does not regard the cost as too great in the circumstances, optimizing moral requirements reappear. Agents who are willing to make sacrifices lose their options, because the options were there in the first place only to limit the costs to them of optimific requirements. The picture is one of a tide of optimizing moral requirements held back by a breakwater protecting the agent's needs (for well-being, for meaningful relationships, and so on). Where the agent's needs are limited or absent, no breakwater is necessary, and the tide floods in with full force. Against the background of this conception, conditional obligations to do the most good possible are readily intelligible and practically inevitable.

5. Half-hearted non-consequentialism and thoroughgoing non-consequentialism

But this conception belongs to a rather half-hearted non-consequentialism, and its account of the moral domain is not compulsory. A more thoroughgoing non-consequentialism rejects the conception, interpreting moral options—what we are morally permitted to do—more straightforwardly. They are simply what we don't infringe any moral duties by doing, where those duties are not arrived at via the characteristically consequentialist procedure of subtracting agent-relative permissions from a general duty to bring about the impartially best outcome (whether or not the permissions include permissions to meet others' claims, as in the expanded version of Horton's view described at the end of section 3 or what Pummer calls 'consequentialism-plus-constraints').²³

My description of Horton's and Pummer's non-consequentialisms as half-hearted may seem tendentious. After all, their views are consistent with a Rossian form of non-consequentialism in which the promotion of good outcomes is only one source of pro tanto moral reasons alongside others. And Ross is standardly taken to be a paradigm example of a non-consequentialist. Moreover, Horton's argument is explicitly contractualist, and contractualism is standardly taken to be a paradigm example of non-consequentialism. If they are only half-heartedly non-consequentialist, then that only goes to show that what's really at issue in the dispute between consequentialists and non-

²³ Pummer, 'Whether and Where to Give', p. 79.

consequentialists is not captured by the distinction between half-hearted and thoroughgoing non-consequentialism.

Since my aim is not to capture what's really at issue in the dispute between consequentialists and non-consequentialists, I have no particular objection to this conclusion. What I want to argue is that an approach to moral theorizing that is characteristic of consequentialism is also exemplified by Pummer's and Horton's arguments, despite their non-consequentialism, and that we should reject it. But it may be useful to say something too about the way in which Rossian and contractualist views can leave room for such characteristically consequentialist thinking.

Ross's non-consequentialism has the character of what Pummer calls 'consequentialism-plus-constraints', or the expanded version of Horton's view that I envisaged at the end of section 3.²⁴ For Ross seems to accept that when we face a situation in which no *prima facie* duty of fidelity, reparation, gratitude, or non-maleficence is applicable, the moral status of our acts is determined by the balance of the remaining three *prima facie* duties in his system, all of which he takes to be instances of the "general principle that we should produce as much good as possible".²⁵ And it is hard to see that principle as anything other than consequentialist. (What Ross says about why *prima facie* duties to promote the good are not themselves consequentialist is unconvincing.)²⁶ So understood, the Rossian view seems no less half-hearted a

²⁴ Pummer, *Whether and Where*, p. 79.

²⁵ W.D. Ross, *The Right and the Good*, ed. Stratton-Lake (Oxford, 2002), p. 27.

²⁶ See Ross, *The Right and the Good*, p. 47.

non-consequentialism than I claim Pummer's and Horton's views are, in the sense that all three seem to embrace the idea that optimific action is the morally required default in the absence of special agent-relative conditions. One significant result of this, as we will see, is that there is no room for permissible non-optimific action in the space left by the absence of others' claims and other special agent-relative conditions.

There may be room for the Rossian to deny this; for example, because what's morally required is not directly a function of the balance of prima facie duties, or because the general principle of good-promotion is not in fact pervasively applicable. But if so, then the Rossian view isn't consistent with the principles advocated by Pummer and Horton after all, and so my description of the latter two as only half-hearted non-consequentialists doesn't commit me to thinking of Ross in the same way.

Unlike the Rossian view, contractualism does not explicitly include any pro tanto duty to promote the good.²⁷ Nevertheless, it does make morality pervasively sensitive to the costs of agents' acts in the sense that everything any agent does is a candidate for moral assessment via the comparison of individuals' complaints arising under principles allowing her to do it with complaints arising under principles forbidding her to do it. If she is permitted to ϕ , that will normally be because being forbidden to ϕ gives rise to a complaint on her part that is greater than any arising under principles

²⁷ I take Scanlonian contractualism as the central case here. See T.M. Scanlon, *What We Owe to Each Other* (Cambridge, MA.: Harvard University Press, 1998).

permitting her to ϕ .²⁸ Therefore, only when the impact on the agent of being required to improve others' situation is sufficiently great is she permitted not to improve their situation, in the sense of bringing them closer to the position they would be in under principles that do not apply only because she can reasonably reject them. The result is that action to improve the situation of others is the morally required default in the absence of special agent-relative conditions such as sacrifices to self and associates, the claims of others, and so on.

There may be room for the contractualist to deny this. One fairly radical way to do so would be to deny that every burden can give rise to a complaint that functions as an input into the contractualist mechanism. For instance, perhaps the pain that one can expect to feel under some set of principles does not furnish one with a complaint under those principles if the pain falls below some threshold. Or perhaps there is a way to exclude some area of action from the scope of contractualist principles from the outset. Alternatively, and less radically, the contractualist may accept that the morally required default is action to improve the situation of others, but then appeal to a very expansive notion of agential autonomy as an agent-relative condition that can ground powerful complaints in the contractualist theory. Appeal to such autonomy wouldn't be the sort of thing an agent could be taken to have forgone as a

²⁸ In some cases it may be that an agent's being forbidden to ϕ gives rise to a complaint on the part of a third party (because the agent's not ϕ -ing generates a cost for that third party, for instance). But these are probably not central cases, and in any case I think the argument I will make can be applied to them too, *mutatis mutandis*.

justification for failing to benefit efficiently merely by acting in any other-benefiting way, unlike appeal to the sacrifice of your arms in *Collapsing Building*. But if any of these strategies is adopted, then contractualism won't be consistent with Horton's view after all (it won't license any inference to a conditional obligation to maximize impartial good or minimize burdens), and so my description of his view as only half-heartedly non-consequentialist won't commit me to thinking of contractualists in the same way.

The description of Pummer's and Horton's arguments as only half-heartedly non-consequentialist, then, seems apt. The basic idea of a more thoroughgoing non-consequentialism, by contrast, is that in moral reasoning there is no default presumption of an optimizing or complaint-minimizing standard, departures from which require special explanation. So there is no pro tanto moral duty to benefit others or promote the good more generally. What we are morally required to do will more typically be conceived as a matter of the limits on the ways in which we live our lives that an adequate appreciation of others' moral standing forces us to recognize.

Versions of the sort of thoroughgoing non-consequentialism I have in mind appear to be philosophically viable alternatives to the more half-hearted approach, and are found attractive by many. Although I will provide some more concrete examples of the thoroughgoing non-consequentialist approach by way of illustration, I will not try to defend any particular view, since my focus is on the nature of conditional obligations and the inference to effective altruism rather than on arbitrating among first-order theories.

6. A thoroughgoing non-consequentialist analysis of rescue cases

It may seem that thoroughgoing non-consequentialism cannot account for intuitively compelling judgments about rescue cases. To see why one might think this, consider *Collapsing Building* again. (I focus on *Collapsing Building* because the judgment about *Arm Donor* is more controversial; as we will see, it is a virtue of the thoroughgoing non-consequentialist analysis that it can make sense of the controversy.) As Horton says, it is surely true that you may either do nothing (on the grounds of the cost to you) or save both children, but you may not save only one child. Yet since the thoroughgoing non-consequentialist rejects the view that the option to do nothing is a permissible deviation from a general requirement of optimific action, it may seem that she is unable to account for the presence of a conditional obligation here. Doesn't the presence of the permission imply the absence of any duties, and the absence of any duties imply that no one is wronged even if you save only the one child?²⁹

This appearance is deceptive, however, for thoroughgoing non-consequentialists can account for the intuitive judgment. Indeed, the explanation can begin in the same way that Horton's does. In some situations, an agent who fails to provide a benefit to others can reasonably appeal to the cost of providing that benefit to justify her failure to provide it. But if she was willing to bear the cost in question for the sake of providing the benefit, then it is not reasonable to appeal to it as a justification for the non-provision. So, in

²⁹ See e.g. Kagan, *The Limits of Morality*, p. 240; Tina Rulli, 'Conditional Obligations'.

these situations, there can be an obligation to provide the benefit, conditional on the willingness to bear the cost.

As we saw in the preceding section, however, in generalizing to *Optimific Altruism* and the conditional obligation of effective altruism, Horton seems to assume that

- (i) *any* failure to provide a benefit to another person occasions a demand for justification,

and that

- (ii) only considerations of cost or agent-relative reasons are sufficient to meet the demand.

These assumptions may be rejected by the thoroughgoing non-consequentialist, and this makes a difference to the way in which conditional obligations are to be understood on the thoroughgoing non-consequentialist account.³⁰

Consider a view according to which morality is fundamentally structured by duties. Some of those duties, such as the duty not to assault others, take the form of more or less permanent prohibitions of certain acts, while others, such as the duties arising from promises and duties of rescue, are activated only in specified situations. The duties are specified in relatively general terms, so that they don't discriminate among cases involving different levels of cost in

³⁰ Even if the range of reasons to which one may appeal for failing to bring about the greater benefit were widened so as to include, for example, the claims of others as in the expanded version of Horton's view that I envisaged at the end of section 3, the thoroughgoing non-consequentialist can still reject assumptions (i) and (ii), giving the explanation that I go on to give.

compliance or benefit to others, at least within a fairly wide range. For example, the duty not to assault another is insensitive to the costs and benefits of compliance both to the agent and to those protected by the duty, at least within a wide range: it applies regardless. There is, however, no general duty to bring about the best outcome as judged from an impartial point of view, or to maximize benefits to others.

Now, failure to comply with one's duties triggers a demand for justification, a demand that one account for one's conduct, since it creates a presumption of wrongdoing—that is, of a failure to give others the consideration demanded by their moral standing, as expressed by the claims correlated with the duties. And when such a demand is triggered, an appeal to the unreasonable costs of compliance in the situation in question (although not the expected costs of compliance with the duty more generally) may answer it. But since there is no general duty to maximize benefits to others, it is not the case that any failure to provide a benefit to another person triggers the demand that one account for one's conduct. So proponents of this sort of view may reject assumption (i).

Even if they accept assumption (i)—perhaps because they sympathise with the idea that it's a condition of showing others the respect due to them as moral equals that we can account for our conduct to them in any situation at all—they may reject assumption (ii), the assumption that only considerations of cost or personal reasons are sufficient to meet the demand. For they may argue that in the absence of the sort of claim to aid that is present in the rescue cases, a perfectly good justification for failing to benefit another is that one wanted to do other things and had no duty not to do them (which is

compatible with one's having an imperfect duty that calls for one to perform beneficial acts of the same sort in similar circumstances on at least some occasions). Of course, appealing to liberties to do other things will invite the further question what justifies taking oneself to have them. But, as we saw, thoroughgoing non-consequentialists are not obliged to accept that the only plausible answer to that question must be derived by subtracting agent-relative permissions from a general duty to bring about the impartially best outcome. So there is no reason to expect that a person has a permission not to benefit another only if the costs to her or others or to valuable projects that she is pursuing are too great.

Intuitively compelling examples such as *Collapsing Building* feature agents confronted with situations in which even most thoroughgoing non-consequentialists recognize that there is a duty to rescue. One type of thoroughgoing non-consequentialist, for instance, will distinguish between our duties to contribute to the establishment and maintenance of just 'basic structures'—systems of legal, political, and economic institutions, tasked with securing basic rights and creating the conditions for peaceful cooperation and flourishing, among other things—and our duties of rescue within the framework of such institutions.³¹ These latter duties address cases of the kind of peril to which basic structure institutions cannot effectively respond: cases of emergency that take place at a distance from institutional resources, calling for

³¹ I take the idea of a basic structure from Rawls, of course. See John Rawls, *A Theory of Justice*, revised edition (Cambridge, MA.: Harvard University Press, 1999), pp. 2–3.

immediate action and commitment of resources from those nearby. Even in a perfectly just world, such cases would arise, and so these duties feature in the thoroughgoing non-consequentialist view, specified in such a way as not to be predictably excessively demanding for agents under the normal background conditions assumed by ideal theory (so they do not call for agents to spend all their time looking for people to rescue, for instance).³²

On this and any other plausible thoroughgoing non-consequentialist view, then, if you are confronted with a collapsing building in which two children are trapped, or with a runaway trolley heading towards people tied to the track, you must, all other things equal, make the practical matter of saving them your first concern—even if, as it turns out, you are justified in not saving them

³² This appeal to the division of moral labour is only one of several approaches any of which thoroughgoing non-consequentialists might adopt. Others might appeal to internally justified limits on the deployment of reactive attitudes; to ideals of independence or autonomous agency; to the significance of relative location; or to the moral fundamentality of the face-to-face encounter, for example. For related arguments, see Barbara Herman, 'Being Helped and Being Grateful: Imperfect Duties, the Ethics of Possession, and the Unity of Morality', *Journal of Philosophy* (2012), pp. 391–411; Fiona Woollard, *Doing and Allowing Harm* (Oxford: Oxford University Press, 2015), chapter 6; Seana Shiffrin, 'Moral Autonomy and Agent-Centred Options', *Analysis* vol. 51, no. 4 (1991); Samuel Scheffler, 'Egalitarian Liberalism as Moral Pluralism', *Proceedings of the Aristotelian Society Supplementary Volume* 79 (2005); F.M. Kamm, 'Does Distance Matter Morally to the Duty to Rescue?', *Law and Philosophy* vol. 19 (2000); David Wiggins, *Ethics* (London, Penguin, 2006), chapter 9; Larry Temkin, 'Obligations to the Needy', given as the Uehiro Lecture, Oxford, 6 November 2017.

yourself by the fact that doing so would involve too great a sacrifice for you. But the correct description for this kind of situation, according to the thoroughgoing non-consequentialist, is not that the people have no claims to rescue, but that failure to satisfy the claims directly can be justified on grounds of cost to the agent. (In the same way, the correct description of a situation in which it has become unreasonably costly to me to keep my promise to you is not that you have no claim to my performance of the promise, but that my failure to satisfy the claim can be justified on grounds of its cost to me.)

Thus the failure to satisfy the claims triggers a demand that the agent account for her conduct, since it is a *prima facie* wrong to the victims. And that demand is not met by appealing to some consideration that would normally justify such a failure if the agent's actions imply that she has discounted the consideration as a reason for failing to satisfy such claims in those circumstances, as they do if she was willing to bear the cost in question for the sake of rescuing anyone. This explains why if you're willing to allow your arms to be crushed in order to save one of two children, then you must also allow your arms to be crushed in order to save both. To do otherwise would gravely wrong the second child, even though you would have been justified in not saving either child had your reason for not doing so been the great cost to you of losing both your arms.

So thoroughgoing non-consequentialist reasoning can explain why there are conditional obligations in *Collapsing Building* and other examples with the same structure. In the case of *Arm Donor*, non-consequentialists may be divided: not everyone thinks we have a duty to save the greater number in this

sort of dilemma, as Pummer recognizes.³³ But either way, the duty to save anyone at all will be understood as a form of the non-consequentialist duty of rescue. The crucial point is that the appeal to duties of rescue does not license any inference to a general principle such as *Avoid Gratuitous Worseness* or *Optimific Altruism*. The children in the collapsing building and the people on the tracks have prima facie claims to your aid that give rise to a demand for justification if you make no effort to rescue them. But since there is no general duty to maximize benefits to others, no one has a claim to benefits beyond those specified by the other duties that a person has, according to thoroughgoing non-consequentialism. And so, as we have seen, in the absence of others' claims, a demand for justification is not triggered even by a failure to provide greater benefits to others that it would be no costlier to the agent to provide than the benefits that she does in fact provide. Or, if it is triggered, a satisfactory justification can be simply that no one had a claim the greater benefits, so that the agent had no duty to provide them. In short, the thoroughgoing non-consequentialist distinguishes between the realm of *claims*, a failure to meet which may be justified on grounds of agential sacrifice, and the realm of benefits *beyond* claims. If others have no claim to our aid in the first place, then they have no claim to more efficient aid than we actually provide either.

³³ 'Whether and Where to Give', p. 84. I say more about this below.

7. The superiority of thoroughgoing non-consequentialism

I believe that the thoroughgoing non-consequentialist approach is more plausible than the half-hearted non-consequentialist alternative. One reason for this derives from the way its account of conditional obligations allows us to preserve intuitively central aspects of the structure of morality.

Conditional obligations in general are a familiar feature of morality, as the example of gratitude shows. But Pummer, Horton, and others working on this topic focus on a particular kind of conditional obligation to benefit others efficiently if one is going to benefit them at all, and, as they have noted, this kind of conditional obligation has various counter-intuitive implications. First, such conditional obligations imply that in some cases a person acts wrongly by saving an innocent person rather than saving no one, given the possibility of saving more people at the same cost.³⁴ Moreover, if the option of saving the greater number at the same cost had not been available, it would have been not only permissible but *heroic* to save the one person, given the sacrifice involved.³⁵ Second, they may imply that someone who makes a great sacrifice in order to rescue others does not do something supererogatory even when she rescues as many as there are to rescue. For given that she rescues anyone, she is *required* to rescue as many as there are to rescue at that sacrifice.³⁶ Third, these

³⁴ Jeff McMahan, 'Doing Good and Doing the Best', in Paul Woodruff [ed.], *Philanthropy and Philosophy: Putting Theory into Practice* [New York: Oxford University Press, forthcoming], section 4.

³⁵ Pummer, 'Whether and Where to Give', pp. 86–7.

³⁶ Horton, 'The All or Nothing Problem', pp. 100–101.

conditional obligations may generate perverse incentives not to do even *some* good in order to avoid being on the hook to do more.³⁷

Some defenders of the conditional obligations in question conclude in light of these points that we must simply rethink standard conceptions of supererogation.³⁸ For instance, Horton suggests that we should give up the idea that anyone ever performs supererogatory acts.³⁹ But if the thoroughgoing non-consequentialist account of the rescue cases does not generate a need for such radical revision to our moral thinking, that counts in favor of thoroughgoing non-consequentialism.

And so it proves. The half-hearted non-consequentialist defenders of the conditional obligation of effective altruism regard the supererogatory as an undifferentiated realm of optimific requirement: when the agent-relative permission not to optimize does not apply, all that's left is a duty to benefit others or do good more generally. By contrast, as we saw in the preceding section, the thoroughgoing non-consequentialist analysis preserves a distinction between the realm of claims and the realm of benefits beyond claims. This latter is the primary realm of the supererogatory. And although, for the realm of claims, the thoroughgoing non-consequentialist must accept some of the counterintuitive implications of conditional obligations that I cited two

³⁷ Pummer, 'Whether and Where to Give', p. 89; Rulli, 'Conditional Obligations'; Horton argues that his view does not have this particular implication at 'The All or Nothing Problem', pp. 97–98.

³⁸ Wessels, 'Beyond the Call of Duty'; Horton, 'The All or Nothing Problem', pp. 100–102; cf. Dorsey, *The Limits of Moral Authority*, chapter 4.

³⁹ Horton, 'The All or Nothing Problem', p. 101.

paragraphs ago,⁴⁰ this does not generate any need to radically rethink or give up the supererogatory. For in the realm beyond claims, it really can be heroic to make great sacrifices to benefit others, even when one might have benefited even more. And no one becomes morally obligated to do more good simply by doing some good. Thoroughgoing non-consequentialism accords with commonsense morality here in a way that half-hearted non-consequentialism does not, even as it returns the intuitively correct judgments about the rescue cases.

There are also other reasons to favor the thoroughgoing non-consequentialist account, which relate to its analysis of rescue cases in particular. To illustrate, consider an important contrast between *Arm Donor* and *Collapsing Building*. In *Collapsing Building*, what's at issue is whether, given that you are going to save a specified person, you should *also* save a second person alongside the first. In *Arm Donor*, what's at issue is whether, given that you are going to save *somebody*, you should save one person or, instead, some other people.⁴¹

⁴⁰ She may not have to accept all of them. By analysing rescue cases in terms of claims, she leaves open the possibility that saving only one person in situations with the structure of *Collapsing Building* is wrongful, because it involves an unjustified failure to meet a claim, but not necessarily wrong. (McMahan makes a related suggestion at 'Doing Good and Doing the Best', section 6.) I cannot develop this idea here, however.

⁴¹ McMahan makes the same point about a different pair of examples in 'Doing Good and Doing the Best', section 2.

As Pummer notes, some philosophers are skeptical about a requirement to save the greater number in examples like *Arm Donor*. They are skeptical typically because they do not think that the right way to think about such examples is by counting up the lives saved on each side of the dilemma and then opting for the larger number.⁴² This doesn't do justice, they tend to think, to the 'separateness of persons': the fact that there is no perspective from which the experiences of fear and death of those who are not saved are all experienced, and no life in which all the subsequent lives of those who are saved are lived.

The separateness of persons isn't a reason to be skeptical about a conditional obligation to save both children in *Collapsing Building*, however, because saving the second child does not come at the expense of saving anyone else. It is a virtue of the thoroughgoing non-consequentialism that I have been describing that it can account for this difference between *Arm Donor* and *Collapsing Building*. It can do so because it recognizes that *Arm Donor* describes a situation in which claims to rescue *conflict*, whereas *Collapsing Building* does not. Even those who are not skeptical about saving the greater number should recognize the significance of this distinction. *Arm Donor* is a tragic case, even if you act heroically and save the one hundred, since someone

⁴² Two classic articles on this issue are Elizabeth Anscombe, 'Who is Wronged? Philippa Foot and "Double Effect"', *The Oxford Review*, vol. 5 (1967), pp. 16-17; and John Taurek, 'Should the Numbers Count?', *Philosophy & Public Affairs* vol. 6, no. 4 (1977), pp. 293-316. Horton points out that *Optimific Altruism* is compatible with scepticism about saving the greater number at 'The All of Nothing Problem', p. 100.

has to die. *Collapsing Building* is not in the same way tragic if you act heroically, since in that case, everyone is saved.

By contrast, the characteristically consequentialist outlook shared by Pummer and Horton risks making it seem as if there is no significant moral difference between these two cases (sheer numbers aside). Pummer even suggests at one point that we can “substitute different specific details [in *Arm Donor*] such that stopping the train on track A promotes one unit of good whereas stopping the train on track B promotes one hundred units of good”, without substantially altering his argument for the conditional obligation of effective altruism.⁴³ I take it that a dilemma between promoting one unit of good and promoting one hundred units of good is not supposed to be tragic. It’s not even supposed to be a dilemma.⁴⁴

8. A thoroughgoing non-consequentialist Rescue Argument?

I believe that Horton’s and Pummer’s defenses of the conditional obligation of effective altruism should therefore be rejected. Thoroughgoing non-

⁴³ Pummer, ‘Whether and Where to Give’, p. 84.

⁴⁴ Some of what Horton says in the course of his argument suggests sensitivity to considerations of conflict (see e.g. ‘The All or Nothing Problem’, pp. 99–100). But when he comes to describe the conditional obligation of effective altruism, the only complications for an optimising strategy that he explicitly registers—beside those relating to the agent’s own special projects and relationships—relate to considerations of incomparability arising from differences in types of goodness.

consequentialism is independently preferable to the half-hearted alternative upon which they rely, and it also provides a more compelling analysis of the rescue cases. But this is not enough to refute the Rescue Argument, since a successful thoroughgoing non-consequentialist version of that argument might still be possible.

The idea would be that even if no inference to a general principle such as *Optimific Altruism* is licensed by the appeal to duties of rescue, the thoroughgoing non-consequentialist who makes that appeal must grant the existence of duties of rescue owed to the world's most needy, who are in the kind of peril that generates claims in rescue cases.⁴⁵ And that would seem to give advocates of efficient charitable giving a way to refurbish the Rescue Argument without falling foul of the objections I have made.

According to the refurbished argument, failure to do what one can to meet the claims of the world's most needy triggers a demand for justification, since it is a *prima facie* wrong to the victims. And although that demand may sometimes be met by appeal to the justifying factor of cost to the agent, it is not met if that factor turns out not to have been operative, as it was not if the agent was willing to bear the cost in question for the sake of meeting any such claims at all. Now, if you donate any money to an inefficient charity whose business is meeting the claims of the world's most needy, that is evidence that the justifying factor of the cost to you of that particular donation, at least, is inoperative. So it cannot be appealed to in justifying the failure to contribute to meeting more such claims at the same cost by donating the same amount to a

⁴⁵ Kagan mounts roughly this objection at *The Limits of Morality*, p. 230.

more efficient charity instead. It follows that you have an obligation to donate to efficient charities whose business is meeting the claims of the world's most needy, conditional upon your donating to any charity in that business.

Clearly, the success of a refurbished Rescue Argument of this sort turns on the crucial premise granting the existence of duties of rescue owed by individuals to anyone in the kind of peril that features in the uncontroversial rescue cases. Must a thoroughgoing non-consequentialist accept that our duties of rescue are so far-reaching? That depends on questions internal to particular thoroughgoing non-consequentialist theories. But it is certainly not clear that any such theory must, if it is to be plausible, affirm such far-reaching duties if rescue on the part of individual agents. Perhaps no plausible theory can disregard the plight of the world's most needy, but even so it is not obvious that addressing that plight entails assigning duties of rescue to individuals. For example, some views, such as the division-of-labor account I sketched earlier, might address that plight via institutional duties that do not straightforwardly translate into individual duties of rescue in the necessary way. Such an account might well conceive their plight as a problem of structural injustice first and foremost, the responsibility of institutional and group agents, and so conclude that each of us is under a duty to contribute to reforming or resisting these unjust institutions, rather than to contribute to the sorts of projects pursued by effective altruists' recommended charities.

Moreover, even if the refurbished Rescue Argument did succeed, it would be in two ways significantly more limited than effective altruists might hope. First, since for almost any given donation size it will not be possible to satisfy every

one of the claims of the world's needy, there will be conflict between those claims. And it is certain that the most efficient available use of any such donation will not satisfy precisely the same claims as every less efficient available use. For the reasons having to do with the separateness of persons that I highlighted in section 7, some thoroughgoing non-consequentialists may therefore deny that in such cases the morally required response is to satisfy as many claims as possible at a given cost to the agent.

Second, even a successful thoroughgoing non-consequentialist version of the Rescue Argument would not establish a general moral obligation of efficiency conditional on any morally optional donation, as *Avoid Gratuitous Worseness* and *Optimific Altruism* purport to do. The obligation established by the refurbished argument would be conditional only on action or willingness to meet claims; that is, only if the agent is subject to duties to needy others that she is nevertheless justified in not fulfilling on grounds of cost to herself. Even a non-consequentialist who grants the existence of extensive claims to be rescued possessed by those in peril will suppose that it is also possible to provide benefits to others that go beyond the satisfaction of any such claims. Using one's resources to provide such benefits will not, on the thoroughgoing non-consequentialist view, give rise to any obligation of efficiency.

9. Conclusion

The conditional obligation of effective altruism defended by Pummer and Horton depends upon a half-hearted kind of non-consequentialism. I have

argued that we should reject such half-hearted non-consequentialism in favor of a more thoroughgoing variety, which preserves the realm of benefits beyond claims and offers a more compelling analysis of rescue cases.

Of course, those consequentialists who see no need to limit the ambitions of effective altruism by allowing for moral options not to benefit the needy in the first place will be unmoved by my objections. And I have not argued against effective altruists who endorse a moral requirement to donate some specified proportion of one's resources to efficient charities. So my discussion does not cast much doubt on effective altruism more generally.

As I noted at the outset, however, the less radical interpretation of effective altruists' flagship claims that is offered by the conditional obligation of effective altruism has both practical and theoretical attractions. So effective altruists might well wish to find a thoroughgoing non-consequentialist way to avail themselves of the Rescue Argument. If I am right, this will be more challenging than they might have anticipated, and the argument's conclusions more limited than they might have hoped.