

Moral Offsetting*

Thomas Foerster

1 Introduction

Consider the following case:

Meat Eating: In the past, Jeff didn't eat meat, since he was concerned with the harm that meat production does to animals. But then he did some calculations, and figured that the harm that he would do to animals by eating meat for a year equals the harm to animals that he could prevent by donating \$200 to animal welfare charities. And he would much rather donate \$200 more to charity and eat meat than neither make the extra donation nor eat meat. Given this, Jeff now eats meat, but each year donates \$200 more than he otherwise would to animal welfare charities.¹

There are a few ways of interpreting Jeff's actions in *Meat Eating*, but here is one natural interpretation. Jeff's donation is supposed to somehow morally offset his consumption of meat. He does something bad (eating meat)² but offsets it by doing something good (making the donation), and ends up morally neutral relative to where he started. Or at least, that is his hope. This can be seen as roughly analogous to carbon offsetting. With carbon offsetting, you do something that increases carbon emissions (such as taking a flight) but offset it by doing something that decreases carbon emissions (such as giving money to promote renewable energy sources), and end up carbon neutral relative to where you started.³

*Forthcoming in *The Philosophical Quarterly*. Please cite the published version: <https://doi.org/10.1093/pq/pqy068>.

¹This case is inspired by Kaufman (2015).

²Throughout this paper I will assume that eating meat harms animals, and that eating meat (without offsetting it, at least) is wrong and blameworthy. This is simply because *Meat Eating* is a useful example to refer to when discussing moral offsetting, and it is a potential instance of moral offsetting only if these assumptions are true. But none of the general points about moral offsetting that I will make in this paper rely on these assumptions.

³Others have drawn an analogy between carbon offsetting and moral offsetting, often with the aim of showing that carbon offsetting is morally objectionable (since, they assume, moral offsetting is morally objectionable). For example, Kinsley (2007) compares carbon offsets to child-abuse offsets, and the now-defunct website cheatneutral.com parodied carbon offsets by comparing them to offsets for cheating on your partner. (Driver 2014 mentions both of these examples.)

This paper explores the idea of moral offsetting. Section 2 clarifies the idea of moral offsetting, distinguishing between two moral offsetting theses—one which talks about wrongness, and one which talks about blameworthiness. Section 3 considers what is relevant to whether one action can morally offset another one, assuming that moral offsetting is ever possible. Section 4 assesses the moral offsetting theses, focusing on whether they are consistent with the leading moral theories and theories of moral worth. Section 5 considers potential benefits of moral offsetting. Finally, section 6 compares moral offsetting to a related practice that I call ‘moral triaging’.⁴

2 Moral offsetting explained

The basic idea of moral offsetting is that good actions can morally offset bad actions. But what, exactly, about bad actions is supposed to be offset by good actions? There are a couple of possibilities here. One thing that good actions might offset is the *wrongness* of bad actions. If Jeff’s donation offsets the wrongness of his meat eating, then he is not doing something wrong when he eats meat. The other thing that good actions might offset is the *blameworthiness* of bad actions. If Jeff’s donation offsets the blameworthiness of his meat eating, then he is no morally worse, overall, after both eating meat and making the donation than he was before doing neither (though he might still be doing something wrong when he eats meat).

To a rough first approximation, these two types of offsetting can be defined as follows:

Wrongness offsetting: G offsets the *wrongness* of B iff doing B alone is wrong, but doing both B and G is permissible.

Blameworthiness offsetting: G offsets the *blameworthiness* of B iff doing B alone would make you a morally worse person than you would be if you did neither B nor G, but doing both B and G would not make you a morally worse person than you would be if you did neither B nor G.

So we get two offsetting theses, each corresponding to a different type of offsetting:

Wrongness Offsetting Thesis: Wrongness offsetting is possible: possibly, for some actions G and B, G offsets the wrongness of B.

⁴Though there is no in-depth discussion of moral offsetting in the philosophical literature, the idea of moral offsetting is not totally new. MacAskill (2015: 142-3) briefly considers a case like *Meat Eating*. Ord (2015: 123) mentions moral offsetting under the label of ‘intrapersonal moral trade’. Alexander (2015) and Askill (2016) discuss moral offsetting in blog posts. Driver (2014) discusses closely related ideas with a focus on environmental ethics. The idea of moral offsetting sometimes also comes up in popular discussions of carbon offsetting, as I noted in the previous footnote.

Blameworthiness Offsetting Thesis: Blameworthiness offsetting is possible: possibly, for some actions G and B, G offsets the blameworthiness B.

That is the basic idea. But the above definitions, though neat and tidy, are too broad: they end up counting too many cases as instances of moral offsetting. So I will add three restrictions to the definitions. These restrictions apply to both the definition of wrongness offsetting and the definition of blameworthiness offsetting. By stipulation, G morally offsets B only if all of the following conditions are satisfied:

- *Harm Condition:* B is harmful even when done in conjunction with G.
- *Avoidability Condition:* It is possible to do G without doing B.
- *Motivation Condition:* G is done for the sake of offsetting B.

The Harm Condition is to prevent cases like the following from counting as instances of moral offsetting:

Soccer Practice: You drop your child off at soccer practice, and pick her up later.

It would be wrong to drop your child off without picking her up later, but dropping her off and picking her up later is permissible. So on the original definition of wrongness offsetting, picking your child up offsets the wrongness of dropping her off. The original definition of blameworthiness offsetting has the same consequence about blameworthiness. These are bad results. The claim that some actions can morally offset other actions in the way that picking your child up “offsets” dropping her off is obvious and uninteresting. Adding the Harm Condition avoids these results, since dropping your child off does not harm her so long as you pick her up later. In contrast, eating meat harms animals even if you donate to animal welfare charities.^{5,6}

The Avoidability Condition is to prevent cases like the following from counting as instances of moral offsetting:

⁵Here harm can be understood according to the *overall counterfactual comparative account*, which says that ‘a harmful event is an event that makes things go worse for someone, on the whole, than they would have gone if the event had not happened’ (Bradley 2012: 396). Again, I am assuming that eating meat harms animals simply because I want *Meat Eating* to be a potential instance of moral offsetting.

⁶The Harm Condition may be too restrictive, since it has the consequence that the wrongness or blameworthiness of only harmful actions can be offset. But there may be harmless wrong or blameworthy actions, such as secretly breaking a promise. To allow for offsetting the wrongness or blameworthiness of this sort of action, the Harm Condition could be replaced with the *Duty Violation Condition*, which says that G offsets B only if B violates a pro tanto duty even when done in conjunction with G. I find the notion of harm clearer than the notion of pro tanto duty, so I will work with the Harm Condition instead of the Duty Violation Condition in this paper. (Thanks to an anonymous referee for raising this point.)

Rescue: I am blocking the path to five people who can be rescued. The only way to get past me is by stepping on my foot. You step on my foot, and rescue the five people.

It would be wrong to step on my foot without rescuing the five people, but stepping on my foot and rescuing the five people is permissible. So on the original definition of wrongness offsetting, rescuing the five people offsets the wrongness of stepping on my foot. The original definition of blameworthiness offsetting has the same consequence about blameworthiness. But this sort of case is importantly different from cases like *Meat Eating*: you have to step on my foot to rescue the five people, but Jeff does not have to eat meat to donate \$200 extra to charity. *Meat Eating* and other candidate instances of moral offsetting are morally questionable largely because the good action could be done without doing the bad action.

The Motivation Condition is less obviously necessary than the other two restrictions. It is satisfied in *Meat Eating*, and it will be satisfied in any clear-eyed attempt to morally offset. I have included it mostly because your motivation for doing something is relevant to how praiseworthy or blameworthy you are for doing it. This is important for the discussion of the Blameworthiness Offsetting Thesis in section 4.2 below.

To be clear, even the modified definitions of moral offsetting are just supposed to help pin down the idea of moral offsetting. There may be cases that fit even the modified definitions of moral offsetting which are importantly different from cases like *Meat Eating*.⁷ If there are such cases, I do not intend for them to count as instances of moral offsetting.

The above definitions define *total* moral offsetting: on the above definitions, if G offsets the wrongness of B then B is not even a little wrong when done with G, and if G offsets of the blameworthiness of B then you are not even a little morally worse after doing B and G than you would be if you had done neither B nor G. But we can also define *partial* moral offsetting. If G merely partially offsets the wrongness of B, then B is wrong even when done with G—it is just less wrong than it would be if it had been done without G. If G merely partially offsets the blameworthiness of B, then doing B and G makes you a morally worse person than you would be if you had done neither B nor G—it just makes you less morally worse than you would be if you had done B without G. It may often

⁷An anonymous referee mentions the following case:

Double Redirection: A runaway trolley is on a track headed towards one trapped person. You redirect the trolley to a second track which has five trapped people on it. Then you redirect the trolley to a third track which has one trapped person on it.

It would be wrong to simply redirect the trolley from the first to the second track, but redirecting the trolley from the first to the second track and from the second to the third track is arguably (though not obviously) permissible. And the Harm Condition is arguably satisfied, since if you had not redirected the trolley from the first to the second track then you would not have redirected it from the second to the third track, and the person on the third track would have been better off. The Motivation and Avoidability Conditions may also be satisfied. So this might count as an instance of moral offsetting on even the modified definitions of moral offsetting.

be useful to think in terms of partial moral offsetting. But for simplicity, I will be concerned only with total moral offsetting in this paper.

3 What can offset what?

The definitions of moral offsetting in section 2 mostly leave open which actions can morally offset which other actions. Of course, it is controversial whether *any* actions can be morally offset. The offsetting theses might be false. But before assessing the offsetting theses, it is worth considering what their proponents might say about which actions can morally offset which other actions. This is interesting in its own right, but it also highlights some resources that proponents of the offsetting theses have in responding to possible objections to their theses.

Here are four possible restrictions on which actions can morally offset which other actions:

- *Similarity*: It might be thought that G can morally offset B only if the subject matter of G is relevantly similar to the subject matter of B. For example, even if Jeff can morally offset his meat eating by donating to an animal welfare charity, it might be thought that he cannot morally offset it by donating to an anti-malaria charity.
- *Consequences*: It might be thought that G can morally offset B only if the consequences of doing G are at least as good as the consequences of not doing B. For example, it might be thought that Jeff can morally offset his meat eating by donating to charity only if the consequences of donating to charity are at least as good as the consequences of not eating meat. This is what Jeff seems to think in *Meat Eating*.
- *Moral reasons*: It might be thought that G can morally offset B only if the moral reason to do G is at least as strong as the moral reason *not* to do B. It might even be thought that G can morally offset B only if the moral reason to do G is *stronger* than the moral reason not to do B. For example, maybe G can morally offset B only if the moral reason to do G is, say, twice or three times as strong as the moral reason not to do B.
- *Supererogation*: It might be thought that G can morally offset any action only if G is supererogatory—only if doing G goes beyond the call of duty. For example, not murdering people is morally good but also morally required, and so it might be thought that Jeff cannot morally offset his meat eating by not murdering people. This might be true even though we have much stronger moral reason not to murder people than to not eat meat.

None of these restrictions is built into the definition of moral offsetting. They are all substantive claims about moral offsetting, each of which might be false even if the offsetting theses are true. Note that a given restriction might apply to wrongness offsetting and not to blameworthiness offsetting, or vice versa. For

example, maybe G can offset the wrongness of B only if G's subject matter is relevantly similar to B's subject matter, but G can offset the blameworthiness of B even if G's subject matter differs significantly from B's subject matter.

As the above list should make clear, a proponent of the offsetting theses does not need to think that only the consequences of B and G are relevant to whether G can morally offset B. For example, suppose that a proponent of the offsetting theses both accepts the moral reasons restriction and thinks that harming is worse than merely allowing harm. Then she might think that even if the consequences of not doing B are as good as the consequences of doing G, G cannot morally offset B. For example, if eating meat harms animals and failing to donate \$200 to an animal welfare charity merely allows animals to be harmed, then we might have stronger moral reason not to eat meat than to make the \$200 donation, even if the consequences of the two are equally good.

4 Assessing the theses

4.1 Wrongness offsetting

In this subsection I will consider whether the Wrongness Offsetting Thesis is consistent with various moral theories. Here I take no stand on whether any of the moral theories that I consider is true. Again, the Wrongness Offsetting Thesis says that wrongness offsetting is possible: possibly, for some actions G and B, G offsets the wrongness of B.

4.1.1 Maximizing act consequentialism

Maximizing act consequentialism (MAC) comes in possibilist and actualist versions.⁸ According to *possibilist* MAC, an action is permissible iff it is part of the best set of actions that you *could* perform, where one set of actions is better than another if its performance would do more good than the performance of the other set. According to *actualist* MAC, an action is permissible iff what *would* happen if you were to perform the action is at least as good as what would happen if you were to not perform the action.⁹

Possibilist MAC is inconsistent with the Wrongness Offsetting Thesis. If an action does unnecessary harm, then it is not part of the best set of actions that you could perform, and so, according to possibilist MAC, is wrong. For example, Jeff *could* be vegetarian but donate extra money to charity anyway, which would do more good than eating meat and donating the extra money to charity. So eating meat is not part of the best set of actions that he could perform, whether or not he donates extra money to charity. So according to possibilist MAC, Jeff's donation does not offset the wrongness of his meat eating.

⁸The labels 'actualism' and 'possibilism' were introduced by Jackson and Pargetter (1986). There is also "hybridism" (see Timmerman and Cohen 2016), but for simplicity I will consider only the actualist and possibilist versions of MAC here. Thanks to an anonymous referee for pointing out the relevance of this distinction to the Wrongness Offsetting Thesis.

⁹My formulations of these views are based on Timmerman and Cohen (2016: 673).

In contrast, actualist MAC is consistent with the Wrongness Offsetting Thesis. Although Jeff could donate extra money to charity even if he were vegetarian, this is not what *would* happen. If Jeff were vegetarian, then he would simply donate less money to charity. The resulting state of affairs is no better than the state of affairs in which he eats meat and donates more money to charity. So according to actualist MAC, Jeff's eating meat is permissible, so long as he would in fact make the extra donation iff he were to eat meat.

More generally, actualist MAC has the following consequence. Doing an unnecessarily harmful action B is permissible if done in conjunction with a good action G, provided that (i) the goodness of the consequences of G at least equals badness of the consequences of B and (ii) you would not have done G if you had not done B. This sort of consequence is sometimes taken as a reason to reject actualist moral theories.¹⁰

4.1.2 Satisficing consequentialism

The basic idea behind all versions of satisficing consequentialism is that you are morally required to do *enough* good, but it is sometimes permissible for you to fail to maximize the good. There are local and global versions of satisficing consequentialism. *Local* versions look at each individual action that you do, and claim that it is permissible so long as it does enough good. *Global* versions look at the whole set of actions that you do throughout your life, and claim that all actions in the set are permissible so long as, together, they do enough good.¹¹

Global satisficing consequentialism (GSC) is consistent with the Wrongness Offsetting Thesis. Suppose that doing B alone would make it the case that you do not do enough good through your life. Then, according to GSC, doing B alone is wrong. But if you do B together with some other action, G, which does a lot of good, then that might make it the case that you *do* do enough good throughout your life. If so, then GSC entails that doing B together with G is permissible.

In contrast, local satisficing consequentialism (LSC) is inconsistent with the Wrongness Offsetting Thesis. There are many versions of LSC, but they all say that whether an action is permissible depends solely on how much good it alone does. One version says that an action is permissible iff it does at least x units of good, for some x . Another version says that an action is permissible iff it does at least $x\%$ of the maximum good that you could do on that occasion, for some x .¹² In any case, the fact that someone does one action, G, which does a lot of good, is irrelevant to whether some *different* action, B, does enough good. For example, if B fails to do x units of good and so, according to the one version of

¹⁰For example, see Carlson's (1997: 263-4) discussion of a case involving a malicious doctor.

¹¹As far as I know, the distinction between local and global satisficing consequentialism is my own. Bradley (2006) considers six versions of satisficing consequentialism, but all are what I have called local versions. I know of no explicit statement of global satisficing consequentialism in the consequentialist literature, but sometimes those who start by explicitly stating a version of local satisficing consequentialism slip into discussing global satisficing consequentialism. For example, Jamieson and Elliot (2009: 243) seem to make this mistake.

¹²Bradley (2006) discusses these and other versions of LSC.

LCS, is wrong, then it *still* fails to do x units of good and is wrong even if it is done in conjunction with some different action that does a lot of good.

4.1.3 Rule consequentialism

There are many versions of rule consequentialism. To save space, here I will consider only the compliance and acceptance versions.¹³

According to *compliance* rule consequentialism, an action is permissible iff it is in accordance with the set of rules that would do the most good if everyone or most people complied with it. Compliance rule consequentialism is arguably inconsistent with the Wrongness Offsetting Thesis. Consider two sets of rules: one which allows moral offsetting, and one which forbids moral offsetting—requiring instead that you *always* do the good “offsetting” action without doing the bad action that it might offset. For example, the rules would require that you be vegetarian, but would also require that you donate money to charity anyway. It seems plausible that the second set of rules would do more good if complied with.

According to *acceptance* rule consequentialism, an action is permissible iff it is in accordance with the set of rules that would do the most good if everyone or most people accepted it. Acceptance rule consequentialism may be consistent with the Wrongness Offsetting Thesis. A set of rules that is less demanding and allows moral offsetting might have a higher compliance rate if accepted than a set of rules that is more demanding and forbids moral offsetting. As a result, the first set of rules might do more good than the second set if accepted.¹⁴

This alone does not show that a set of rules that allows moral offsetting would do the *most* good if accepted. After all, there are other ways for a set of rules to achieve a high compliance rate. For example, the rules of commonsense morality are undemanding enough to have a high compliance rate, but they arguably do not allow moral offsetting. Nonetheless, this at least shows that acceptance rule consequentialism is not obviously inconsistent with the Wrongness Offsetting Thesis.

4.1.4 Kantian ethics

According to Kantian ethics, an action is permissible iff it is in accordance with the categorical imperative. Kant formulates the categorical imperative in a few different ways, but here I will focus on the formula of humanity, which says that you should always treat each person as an end, never merely as a means.¹⁵ Kantian ethics is arguably inconsistent with the Wrongness Offsetting Thesis. The fact that someone does one action, G, which is morally good, is irrelevant to whether a *different* action, B, treats someone as an end. If B fails to treat someone as an end and so, according to Kantian ethics, is wrong, then it *still*

¹³See Hooker (2016) for a discussion of various versions of rule consequentialism, including the compliance and acceptance versions.

¹⁴Thanks to an anonymous referee for raising this point.

¹⁵Kant (2012: 49).

fails to treat someone as an end and is wrong even if it is done in conjunction with some different action that is morally good.¹⁶

Kantian ethics does, however, allow for cases that are structurally similar to wrongness offsetting. This is because it allows for “imperfect” duties, in addition to “perfect” duties. Roughly, perfect duties are things that you must always do, whereas imperfect duties are things that you must do at least sometimes, but you do not have to do all the time. For example, Kant thinks that we have a perfect duty not to lie but an imperfect duty to help others. This means that we must never lie but we only sometimes have to help others. If you have an imperfect duty to help others, then whether you act wrongly by failing to help someone on a given occasion depends on whether you have helped others on *different* occasions.

To see how imperfect duties give rise to cases that are structurally similar to wrongness offsetting, let G be helping Gail and B be not helping Ben. Suppose that, since you have an imperfect duty to help others, you are morally required to help either Gail or Ben, but you are not morally required to help both of them. Then doing B alone is wrong, but doing both B and G is permissible.¹⁷ Given the Harm Condition, whether this counts as an instance of wrongness offsetting depends on whether failing to help Ben counts as harming him. Failing to help does count as harming on some accounts of harm¹⁸, but this consequence is sometimes taken as a reason to reject these accounts of harm.

In any case, this opens the door to at most a limited range of actions whose wrongness can be offset: actions that involve failing to do imperfect duties. The wrongness of actions that involve failing to do *perfect* duties cannot be offset, since you must always do what you have a perfect duty to do. So even if Kantian ethics allows for some wrongness offsetting, it does not allow for offsetting the wrongness of, say, intentionally harming others, since we have a perfect duty not to intentionally harm others.

4.2 Blameworthiness offsetting

Again, the Blameworthiness Offsetting Thesis says that blameworthiness offsetting is possible: possibly, for some actions G and B, G offsets the blameworthiness of B. The Wrongness Offsetting Thesis entails the Blameworthiness Offsetting

¹⁶What if your actions are directed at the same person? For example, I do you a big favor, and then make some onerous demand of you. Maybe simply making the demand of you would be treating you merely as a means, but both doing you the favor and making the demand of you does not treat you merely as a means. Still, this would arguably not count as an instance of moral offsetting, since making the demand of you does not harm you. In any case, the paradigmatic candidate cases of moral offsetting, such as *Meat Eating*, are ones that concern different groups of people or animals. For example, Jeff harms some animals by eating meat and then helps *different* animals by donating to animal welfare charities. Thanks to Julia Markovits for discussion here.

¹⁷Thanks to Julia Markovits for drawing my attention to this sort of case.

¹⁸Most notably, the counterfactual comparative account of harm, which I am officially working with. See note 5 above.

Thesis.¹⁹ But it is more interesting to consider whether the Blameworthiness Offsetting Thesis is true if the Wrongness Offsetting Thesis is false. So in the rest of this subsection, I will assume for the sake of argument that the Wrongness Offsetting Thesis is false and consider whether, given this assumption, the Blameworthiness Offsetting Thesis is true.

Here is an intuitive but flawed argument for the Blameworthiness Offsetting Thesis. Generally, doing something bad is blameworthy and makes you a morally worse person. Generally, doing something good is praiseworthy and makes you a morally better person. So there should be some pairs of actions, B and G, such that the amount to which doing B makes you a morally *worse* person is less than or equal to the amount to which doing G makes you a morally *better* person. Given this, if you do B alone you will end up a morally worse person, but if you do both B and G you will end up at least as morally good as you were before doing neither B nor G. So G offsets the blameworthiness of B, and the Blameworthiness Offsetting Thesis is true.

But this argument, as I said, is flawed. Specifically, it is invalid. It should be relatively uncontroversial that there are some actions B and G, as above, such that doing both B and G leaves you as morally good as you would be if you had done neither B nor G. The problem is that it does not follow from this that G offsets the blameworthiness of B. Given the Motivation Condition, doing both B and G counts as an instance of moral offsetting only if you do G for the sake of offsetting B. And it might be that when a good action is done with the aim of offsetting a bad action, it does not count as praiseworthy, even if it *would* count as praiseworthy if it were done with some other aim.

There is good reason for proponents of the Blameworthiness Offsetting Thesis to worry about this. According to the leading accounts of moral worth, there is a tight connection between your motivation for doing something and how praiseworthy or blameworthy you are for doing it. For example, according to Markovits's *Coincident Reasons Thesis*, your action is praiseworthy iff, and to the degree that, '[your] motivating reasons for acting coincide with the reasons morally justifying the action' (2010: 205).²⁰ Your motivating reasons for acting are what motivate you to act in the way you did. The reasons morally justifying an action are what make it the case that the action is morally good or justified.

To see how the Coincident Reasons Thesis presents a challenge for the Blameworthiness Offsetting Thesis, think back to Jeff. On the simplest reading of *Meat Eating*, his *motivating reason* for donating to animal welfare charities is simply to offset his meat eating. But the reason *morally justifying* donating is that it helps animals. Since these differ, the Coincident Reasons Thesis entails that Jeff is not praiseworthy for donating. And if he is not praiseworthy for

¹⁹To see this, suppose that G offsets the wrongness of B. Then doing B alone is wrong and so (in normal circumstances) makes you a morally worse person, but doing B together with G is permissible and so does not make you a morally worse person. Then, by the above definition of blameworthiness offsetting, G offsets the blameworthiness of B. It is, however, a bit misleading to say that G offsets the blameworthiness of B in this case, since doing B is not blameworthy at all if done in conjunction with G. This is not a case in which doing B *is* blameworthy, but the praiseworthiness of doing G cancels it out.

²⁰Arpaly (2002) has a similar account of moral worth.

donating, then donating does not make him a morally better person, and so cannot offset the blameworthiness of his meat eating. The same goes for any attempted case of offsetting, provided that you do the good action solely with motivation to offset the bad action.

But despite the above challenge, I think that the Blameworthiness Offsetting Thesis is true even if the Wrongness Offsetting Thesis is false. This is because there are two complications that the previous paragraph glosses over.

The first complication is that there are different ways of understanding ‘the motivation to offset the bad action’. Above I tacitly understood it as the motivation to make it the case that performing the bad action does not leave you, on the whole, morally worse than you were before performing it. Call this ‘the *inward-looking* offsetting motivation’. This motivation clearly differs from the reason morally justifying a good action. But ‘the motivation to offset the bad action’ could instead be understood as the motivation not to leave the world, on the whole, worse than it was before you performed the bad action. Call this ‘the *outward-looking* offsetting motivation’.

Unlike the inward-looking offsetting motivation, the outward-looking offsetting motivation is arguably grounded in concern for things in the world. After all, if you do not care about people or animals or anything else in the world, then it is hard to see why you would care whether your actions leave the world, on the whole, a worse place. So if Jeff has an outward-looking offsetting motivation for donating to animal welfare charities, then he arguably is motivated by at least partly by concern for animals, and so there *is* an overlap between his motivating reasons for donating and the reasons morally justifying donating. If this is right, then the Coincident Reasons Thesis entails that he is praiseworthy for donating.

The second, related complication is that we generally do not do things for just one reason. Jeff might care a bit about the suffering of animals, and that might be part of his motivation for donating to animal welfare charities, even if his *main* motivation for donating is to offset his meat eating in the inward-looking sense. If so, then there is a partial overlap between his motivating reasons for donating and the reasons morally justifying donating. Then Jeff is somewhat praiseworthy for donating, according to the Coincident Reasons Thesis. But he is less praiseworthy than he would be if his motivating reasons coincided more fully with the reasons morally justifying donating. Given this, it is unclear whether the limited praise that Jeff deserves for donating \$200 is enough to outweigh the blame that he deserves for eating meat.

But what if Jeff donates a lot more than \$200 to charity, or does something else that is exceptionally good? For example, what if he donates a kidney to a stranger? Suppose that he is motivated to do these things mostly by the desire to offset his meat eating in the inward-looking sense, but that he is also motivated partly by a desire to help animals or people. Even though he is less praiseworthy for doing these things than he would be if his motivating reasons coincided fully with the morally justifying reasons, it seems plausible that he is still very praiseworthy for doing them. Donating a kidney to a stranger is so exceptionally good that even, say, 5% of the full praise deserved for doing it is a lot of praise. It is presumably enough to outweigh the blameworthiness of eating

meat.

The upshot is that the Blameworthiness Offsetting Thesis is true even if the Wrongness Offsetting Thesis is false. Importantly, though, you succeed in offsetting the blameworthiness of a bad action only if your motivating reasons for performing the good action coincide at least partly with the reasons morally justifying the good action. And the less overlap there is between your motivating reasons and the reasons morally justifying the good action, the more significant the good action has to be in order to fully offset the blameworthiness of the bad action.

5 Benefits of offsetting

In this section I will consider the benefits that moral offsetting would have if the offsetting theses were true. There are two reasons why the benefits are worth considering whether or not the offsetting theses are true. First, I suspect that the benefits partly explain the appeal of the offsetting theses. Second, and more importantly, the benefits point to important insight about how to think about our moral obligations. In the next section, I will explain why this insight is important even if the offsetting theses are false.

Throughout this section I will assume that moral offsetting is not only possible, but also relatively cheap. For example, I will assume that Jeff can morally offset his meat eating by donating \$200 extra to charity, and that he does not have to do anything as extreme as, say, donating a kidney to a stranger. If moral offsetting cost so much, then it would be too expensive to be worthwhile.

Assuming that moral offsetting is both possible and relatively cheap, there are two main benefits of moral offsetting. The first, most obvious benefit is that it allows us to better satisfy our preferences without doing worse morally.²¹ For example, Jeff prefers both eating meat and making the extra donation to neither eating meat nor making the extra donation, and so if he succeeds in offsetting his meat eating then he better satisfies his preferences without doing worse morally.

There are a couple reasons why you might prefer doing both B and G to doing neither B nor G. It might be that doing B increases your well-being more than doing G decreases your well-being. For example, Jeff might think that eating meat makes his life go better—it allows him to enjoy bacon and cheeseburgers—and that donating a bit more to charity does not have much effect on his well-being. Or it might be that refraining from doing B takes more effort than doing G, whether or not this significantly affects your well-being. For example, Jeff might think that it would take a lot of effort to be vegetarian, and it is easier to simply eat meat and donate more to charity.

Importantly, offsetting for either of these reasons increases your *moral efficiency*. For example, suppose that Jeff offsets his meat eating because donating

²¹The phrase ‘doing worse morally’ is deliberately ambiguous in order to cover both wrongness offsetting and blameworthiness offsetting. One way of doing worse morally is to do more wrong things, which wrongness offsetting prevents. The other way of doing worse morally is to become a morally worse person, which blameworthiness offsetting prevents.

\$200 extra to charity lowers his well-being less than eating meat increases his well-being. Then he is doing just as well morally, but has a higher well-being. Or suppose that Jeff offsets his meat eating because it takes him less effort to donate \$200 extra to charity than to be vegetarian. Then he is doing just as well morally, but putting less effort into it. In either case, he is getting more bang (moral accomplishment) for his buck (well-being or effort).²²

It might look like moral efficiency is something that you would care about only for self-interested reasons—it makes *your* life go better, and takes less effort for *you*. But moral efficiency is also *morally* important.

This brings us to the second benefit of moral offsetting: in being more morally efficient, it allows us to achieve more morally before we reach our breaking points. There is only so much well-being that we are willing to sacrifice for morality, and only so much effort that we are willing to put into acting morally. These are our breaking points.²³ If we are more morally efficient, then that leaves us more well-being and effort to reinvest into acting morally. For example, if Jeff finds it easier to offset his meat eating than to be vegetarian, then offsetting leaves him effort to do something good that he would not otherwise have had the effort to do. Of course, he doesn't have to reinvest the saved effort into acting morally, but he at least has the option to do so.

Here is an analogy. Suppose that I am a body builder who is willing to work out for eight hours a day. Now suppose that I develop a more efficient workout routine, which allows me to accomplish in only four hours what had taken me eight hours. In response to this increase in efficiency, I could be lazy and decide to work out for only four hours a day, accomplishing just as much as I did before. This is analogous to someone who fails to reinvest the gains of moral efficiency into acting morally. Alternatively, I could keep on working out for eight hours a day, now at the increased level of efficiency, and accomplish even more than I did before. This is analogous to someone who reinvests the gains of moral efficiency into acting morally. Note that any serious body builder should want to work out efficiently. Efficiency is not just for lazy people.

Again, this is all assuming that the offsetting theses are true. If the offsetting

²²More formally, moral efficiency can be defined as follows (distinguishing efficient use of well-being from efficient use of effort):

Moral efficiency with respect to well-being: The moral efficiency of ϕ -ing equals the strength of the moral reason to ϕ divided by the amount that ϕ -ing lowers your well-being.

Moral efficiency with respect to effort: The moral efficiency of ϕ -ing equals the strength of the moral reason to ϕ divided by the amount of effort it takes you to ϕ .

Strictly speaking, the definitions should be relativized to a specific person, since how morally efficient a given action is might vary from person to person. For example, the amount of effort that it takes me to ϕ might be greater than the amount of effort that it takes you to ϕ . Then, all else being equal, ϕ -ing will be more morally efficient for you than for me. But for simplicity, I will usually not bother relativizing talk of moral efficiency to a specific person.

²³To be clear, by 'breaking point' I just mean the point at which you are not willing to do more for morality. But even if you are at your breaking point in this sense, it will almost always be possible for you to do more for morality. You just do not *want* to do any more.

theses are false, then these apparent gains to moral efficiency will be (mostly)²⁴ illusory. But the idea of moral efficiency is important even if moral offsetting is not a way of being more morally efficient. There might be other ways of being morally efficient. I explore one way in the next section.

6 Moral triaging

Suppose that I wake up one morning and realize that there are seventeen ways in which I am failing morally—that there are seventeen things that I am morally required to do that I am not currently doing. Ideally, I would aim to do all of them. But seventeen is a lot, and knowing what I know about myself, I know that that is an unrealistic goal. I know that before I do all seventeen things, I will likely reach the breaking point of how much well-being I am willing to sacrifice for morality, or how much effort I am willing to put into acting morally.²⁵ So I need to decide which of the seventeen things that I am morally required to do that I will try to do before I reach my breaking point. How should I make this decision?

One way to think about this is to imagine that I am making a to-do list of the things that I am morally required to do. After making the list, I will start by doing the first item on the list, and then work my way to the bottom. Again, knowing what I know about myself, I know that I will likely reach my breaking point before getting through everything on the list. The question is then: how should I order the items on the list?²⁶

A natural answer is that I should order the items by *moral importance*: I should put what I have the most moral reason to do at top, and then put what I have the next most moral reason to do, and so on. A different answer is that I should order the items by *moral efficiency*: I should put what gives me the most moral bang for my buck at top, and then put what gives me the next most moral bang for my buck, and so on. For example, suppose that I have slightly stronger moral reason to do A than B, but doing B is much easier for me than doing A. Then B gives me more moral bang for my buck, and so on this ranking it would be placed ahead of A.

To see why the second ranking is better, consider the following case. I am morally required to do A, B, and C. There is slightly stronger moral reason to do A than to do B or C, but doing A would take a huge amount of effort,

²⁴Even if the offsetting theses are false, there might be some moral benefit to people *attempting* to offset. For example, even if Jeff fails in his attempt to offset his meat eating, he might end up doing morally better overall after attempting to offset his meat eating than before doing so, provided that he reinvests the saved effort into acting morally.

²⁵I am not assuming that it is impossible for me to do all seventeen things. So this case is consistent with the view that ‘ought’ implies ‘can’.

²⁶A complication, which I will set-aside, arises from the fact that some of these things will be time-critical (for example, picking your child up from soccer practice), and so should get pushed up to the top of the list simply on that basis. In contrast, the moral obligations that I primarily have in mind are things like giving money to charity, reducing your carbon footprint, and being vegetarian—things that will still be there for you to do in a day or a year if you do not do them right now.

putting me at my breaking point. Doing B and C would take much less effort, and I could do both of them before reaching my breaking point. According to the first ranking, I should put A at the top of my list, since I have the most moral reason to do it. This will lead me to do A and nothing else. According to the second ranking, I should put B and C at the top of my list. This allows me to do both B and C, which together are much more morally important than A alone. This is a better result.

For a concrete example, think back to Jeff in *Meat Eating*. Suppose that he is morally required both to be vegetarian and to donate some of his money to charity. To make things simple, suppose also that the strength of the moral reason to be vegetarian equals the strength of the moral reason to donate \$200 more to charity. Since it takes less effort for Jeff to make the donation than to be vegetarian but the two are equally morally important, making the donation is more morally efficient than being vegetarian. Given this, the second ranking would have Jeff put making the donation ahead of being vegetarian on his to-do list. Note that Jeff might reach his breaking point before getting to the ‘be vegetarian’ item on his to-do list, and so de-prioritizing being vegetarian might amount to never being vegetarian.

When someone ranks actions in this way, in terms of moral efficiency, we can say that they are engaging in *moral triaging*. Moral triaging is roughly analogous to triaging in a medical context, which doctors might engage in after a mass casualty event. This involves turning away patients who will likely die no matter what, and giving priority to patients who will likely benefit the most from treatment.

Moral triaging might look similar to moral offsetting, but the two are importantly different. Moral triaging is simply a way of prioritizing morally significant actions. It does not aim to make wrong actions permissible, or to stop you from becoming a morally worse person. So unlike moral offsetting, it does not rest on any controversial moral theories or theories of moral worth. When Jeff ranks donating above being vegetarian, it is not because he believes that that makes eating meat permissible, or that it leaves him as morally good as he would be otherwise. He is simply trying to work through his moral obligations in the best way possible.

When you triage, you make a plan that takes into account the fact that you will likely fail to fulfill all of your moral obligations. Opponents of triaging might object that you should always plan to fulfill all of your moral obligations, however unlikely it is that you will succeed in doing so. To defend this claim, they might argue that if it is wrong to do something, then it is also wrong to *plan* to do that thing. But it is wrong to fail to fulfill all of your moral obligations, and when you triage you are planning to fail to fulfill all of your moral obligations. It follows from this that triaging is wrong.

This objection rests on a misunderstanding of triaging. When you triage, you are making a plan that accounts for the *possibility* of your failing to fulfill all of your moral obligations, but you are not planning not to fulfill all of them. You are just minimizing the harm that will come if, as is likely, you fail to fulfill all of them. If it turns out that you are willing to sacrifice more for morality

than you originally thought, your plan does not tell you to stop working your way through your to-do list of moral obligations.

Do most of us have reason to engage in moral triaging? This depends partly on how much morality demands of us. If an extremely demanding moral theory, such as MAC, is true, then pretty much everyone will reach their breaking points before fulfilling all of their moral obligations. In this case, we should triage. If a less demanding moral theory, such as one resembling commonsense morality, is true, then most people will be able to fulfill all of their moral obligations before reaching their breaking points. In this case, triaging is unnecessary, at least as far as our moral obligations go: we can simply do everything we are morally required to do (don't steal, don't murder, and so on) without making any trade-offs.

But even if a less demanding moral theory is true, we might still have reason to engage in moral triaging. For example, on most moral theories that are less demanding than MAC, there are supererogatory actions. We are then faced with the question of which supererogatory actions to do. Of course, you could decide not to do any of them, in which case triaging is of no use to you here. But most of us want to do at least some morally good things that we are not morally required to do. Triage gives us a way of deciding which ones to do: it tells us to start with the most morally efficient supererogatory actions, and then work our way down to the least morally efficient ones.

A real-life example of this comes from discussions of non-directed kidney donation in the effective altruism community (for example, see Kaufman 2012). Some of those who choose not to donate a kidney reason that they could do as much good by donating more money to charity, and that they would rather donate more money than donate a kidney. Given this, they conclude that they should start by donating more money to charity, and then donate a kidney (if ever) only after they get to the point at which they would rather donate a kidney than donate any more money to charity. In other words, donating money is currently more morally efficient for them than kidney donation, so they should prioritize donating money over kidney donation.

7 Conclusion

In this paper, I have discussed wrongness offsetting, blameworthiness offsetting, and moral triaging.

Whether the Wrongness Offsetting Thesis is true depends on very specific details of which moral theory is true. The Wrongness Offsetting Thesis is consistent with some moral theories (actualist MAC, GSC, and arguably acceptance rule consequentialism), but inconsistent with others (possibilist MAC, LSC, Kantian ethics²⁷, and arguably compliance rule consequentialism).

The Blameworthiness Offsetting Thesis is true if the Wrongness Offsetting Thesis is true. The Blameworthiness Offsetting Thesis is true even if the Wrongness Offsetting Thesis is false, but in this case you succeed in offsetting the

²⁷With a possible exception for imperfect duties, discussed above.

blameworthiness of a bad action only if your motivating reasons for performing the good action coincide at least partly with the reasons morally justifying it—you cannot be motivated solely by concern for your own moral goodness.

Finally, moral triaging is simply a way of prioritizing morally significant actions. Unlike wrongness offsetting, the practice of moral triaging is not hostage to any specific moral theories.²⁸

References

- Alexander, S. (2015). Ethics Offsets. Retrieved December 4, 2018, from <http://slatestarcodex.com/2015/01/04/ethics-offsets/>
- Arpaly, N. (2002). Moral Worth. *The Journal of Philosophy*, 99(5), 223–245.
- Askell, A. (2016). Can we offset immorality? Retrieved December 4, 2018, from <http://www.rationalreflection.net/can-we-offset-immorality/>
- Bradley, B. (2006). Against Satisficing Consequentialism. *Utilitas*, 18(2), 97–108.
- Bradley, B. (2012). Doing Away with Harm. *Philosophy and Phenomenological Research*, 85(2), 390–412.
- Carlson, E. (1999). Consequentialism, Alternatives, and Actualism. *Philosophical Studies*, 96(3), 253–268.
- Driver, J. (2014). Moral Bookkeeping, Consequentialism, and Carbon Offsets. In A. Hiller, R. Ilea, & L. Kahn (Eds.), *Consequentialism and Environmental Ethics* (pp. 164–173). Routledge.
- Hooker, B. (2016). Rule Consequentialism. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Winter 2016). Metaphysics Research Lab, Stanford University. Retrieved December 4, 2018, from <https://plato.stanford.edu/archives/win2016/entries/consequentialism-rule/>
- Jackson, F., & Pargetter, R. (1986). Oughts, Options, and Actualism. *Philosophical Review*, 95(2), 233–255.
- Jamieson, D., & Elliot, R. (2009). Progressive Consequentialism. *Philosophical Perspectives*, 23(1), 241–251.
- Kant, I. (2012). *Groundwork of the Metaphysics of Morals*. Cambridge University Press.
- Kaufman, J. (2012). Altruistic Kidney Donation. Retrieved December 4, 2018, from <https://www.jefftk.com/p/altruistic-kidney-donation>
- Kaufman, J. (2015). How Bad Is Dairy? Retrieved December 4, 2018, from <https://www.jefftk.com/p/how-bad-is-dairy>
- Kinsley, M. (2007). Credit for Bad Behavior. *Time*. Retrieved December 4, 2018, from <http://content.time.com/time/magazine/article/0,9171,1635840,00.html>
- MacAskill, W. (2015). *Doing Good Better*. Gotham Books.
- Markovits, J. (2010). Acting for the Right Reasons. *Philosophical Review*, 119(2), 201–242.
- Ord, T. (2015). Moral Trade. *Ethics*, 126(1), 118–138.

²⁸Thanks especially to Julia Markovits, and also to Avi Appel, Derk Pereboom, and two anonymous referees for *The Philosophical Quarterly*.

Timmerman, T., & Cohen, Y. (2016). Moral Obligations: Actualist, Possibilist, or Hybridist? *Australasian Journal of Philosophy*, 94(4), 672–686.