Offsetting Harm*

Mike Deigan
`mike.deigan@rutgers.edu`

I increase the amount of CO_2 in the atmosphere by driving, buying plane tickets, ordering goods to be delivered to me, and so on. And the more CO_2 in the atmosphere, roughly, the worse various climate change-related harms will be. So I worsen or, in other words, increase these harms.[1] Aside from polluting, we have many other opportunities to contribute to processes which result in more or less harm, depending on how we contribute. We can increase demand for factory-farmed animal products, for example, or add to social pile-ons. Given their ubiquity, it is important to know how we can permissibly interact with such processes.

Limiting our attention to cases where we increase some harm by a perceptible amount, we might think that there is no special theoretical difficulty here.[2] Increasing harm, after all, is just a way of doing harm. So we could reasonably hope to build an ethical theory around less complicated, more direct cases of doing harm, then apply this theory to the practically significant cases of merely increasing harm. But I think this is a mistake.

[1]See Broome (2019) for a recent argument against denials of this unfortunate fact.

[2]Much has been said about cases where one contributes to some collective harm in a way that seems to make no perceptible difference (Nefsky 2019). Though there may be overlap in applications, the harm increases I will be discussing are ones that I will assume do make a perceptible difference, and so do not face the same kind of inefficacy problems discussed in the collective/imperceptible harm literature.

There's a way of making some harm increases permissible that is not available in simpler cases of doing harm, one that we are bound to miss if we only consider the simpler cases.

Besides being able to contribute to harmful processes in ways that increase the harms that result, sometimes we can also contribute to them in ways that decrease those harms. We can buy carbon offsets, promote veganism or pay others to not buy animal products, reduce what others add to social harms, and so on. This means that sometimes we can act in ways that increase a harm, but still have total contributions *to that very harm* that are net neutral or negative. Often a harm increase seems permissible if it is *offset* in this way with an equal or larger harm decrease.

This paper is an exploration of offsetting harm increases from the perspective of more-or-less commonsense deontological theory.[3] I will argue that the standard deontological constraint against doing harm cannot accommodate permissible offsetting and so should be replaced by a constraint that can.

1 Zap Offsetting Cases

Let's start with some artificially simple cases.

Innocent *C* will soon get a mildly painful zap from evil *B*'s zapping machine. How much the zap will hurt is determined by how much weight is on a scale attached to the machine: the more weight on the scale, the more painful the zap. One can remove a weight from the scale, but only if one adds a weight of one's own or pays a large fee. There are no other ways

---

[3]More specifically, I'll be assuming there are moderate constraints that sometimes require agents to not do something even if that would lead to a more valuable outcome, as well as options both of agent-sacrificing and agent-promoting kinds, and a morally relevant distinction between doing and allowing. These assumptions are very controversial, but commonly enough held for a project which takes them as starting points to be of interest, even to those like myself who have doubts about them.

2

to interfere with the zapping. Currently there are 7 lbs worth of weights on the scale. The bystander *A* is aware of these facts.

Now consider *A*'s actions in the following four variants of the case.

Zap Non-Interference

A neither adds any weight to the scale nor pays the large fee to remove a weight, leaving the total weight on the scale at 7 lbs.

A's inaction here seems permissible. Sometimes one is obligated to aid a stranger facing some impending harm, but assuming that the fee is sufficiently high, the differences in zap painfulness sufficiently small, and that *A* doesn't have any special obligations to *C*, they would not wrong *C* by not taking on a significant cost to themselves in order to reduce the weight on the scale.

Zap Increase

A has a 1 lb weight that they don't feel like holding onto anymore, so they place it on the scale without removing any of the weights that were already there, leaving the total weight on the scale at 8 lbs.

In Zap Increase, *A*'s act makes the harm to *C* worse. Sometimes increases to harms are permissible, but this one is clearly impermissible.

Zap Offsetting (Negative)

A is carrying a 1 lb weight, but would prefer to have a 2 lb weight. *A* places the 1 lb weight on the scale, and removes a 2 lb weight, leaving the total weight on the scale at 6 lbs.

Any harm increase that comes from the addition of a weight is offset by the removal of a heavier weight. In this case, *A* makes things better for *C* by making the trade. Given that they had no way to make things even better for *C* without taking on significant costs to themselves, this seems permissible.

Is net harm reduction necessary for permissibility? I think not. Neutrality seems to suffice.

> **Zap Offsetting**
>
> *A* has a 1 lb weight that is painted red, but would prefer to have a blue one. In this scenario, there are only 1 lb weights on the scale. *A* places the red weight on the scale and removes a blue one, leaving the total weight on the scale at 7 lbs.

C seems to have no more complaint against *A* here than they do in Zap Non-Interference. It would have been nice of *A* to pay the large fee to make things better for *C*, but this is not morally required.

So it seems that one can permissibly do something that makes a harm worse, such as adding a weight to the zapping machine scale, if one offsets it by reducing the harm by at least as much, such as by removing a different weight from the scale. The question I'm interested in is why increasing harm is permissible when paired with offsetting of this kind.

A straightforward answer seems to be available to those who go in for a familiar deontological constraint against doing harm. It's impermissible to do harm, except in special circumstances. In Zap Increase, *A* does harm to *C* without good enough reason, so acts impermissibly. But in the offsetting cases, *A* doesn't do any harm or violate any other constraints, so acts permissibly, just like in Zap Non-Interference. Simple as that, right?

Wrong. This straightforward account fails, since one *does* do harm in offsetting cases (§2). Moreover, the permissibility of this harm cannot be captured by the standard provisos to the constraint against harm, such as the proviso that one may do harm if it is required for bringing about enough good (§3).

What I take the permissibility of offsetting cases to show is that the distinction that really matters is not one between harmful actions and non-harmful ones, but between actions which involve *unoffset harm increases*

and those that don't. Doing something that increases a harm is not itself even pro tanto wrong, making a total contribution to it that is net positive is. I thus propose we replace a constraint against doing harm with one against making unoffset harm increases (§§4–6).

If this is right, it should be of interest to those concerned with finding the best way to formulate a deontological theory: a cornerstone of that approach needs revision. This revision may require us to rethink the source of deontological constraints (§7). It is also of practical interest. Given that emitting CO_2 will worsen the harms of climate change, is it permissible, as John Broome (2012) argues, to emit CO_2 so long as one reduces the amount of CO_2 in the atmosphere by at least as much?[4] And given that buying animal products can increase demand for them, and so increase the amount of suffering animals, would it be permissible to purchase such products so long as one, say, pays other people not to buy as much as they would have otherwise?[5] Whether such behavior is permissible will depend on what makes offsetting permissible, and whether these cases have the relevant features (§8).

2 Offsetting Not Prevention

Let us return to the natural proposal about the permissibility of offsetting that I say we should reject: offsetting cases are permissible because they don't really involve doing harm, since what offsetting does is prevent an

---

[4]As Broome notes, this is not as simple as having trees planted, since once trees die they typically decompose and return the carbon they've stored back into the atmosphere. There are ways, though, to artificially extract and store carbon more or less permanently, even if they are currently rather expensive expensive (Herzog (2018, Ch. 6)). Broome himself recommends "preventative offsetting", which doesn't take carbon from the atmosphere, but rather prevents gas that would have been emitted from being emitted (Broome 2012, p. 87). I am somewhat skeptical that this prevention is sufficient for offsetting in the sense I will be discussing, but leave discussion of this issue for elsewhere.

[5]The question of meat offsetting has been raised by MacAskill (2015, Ch. 8), who finds carbon offsetting permissible, but not adultery or meat offsetting.

action that would have been harmful from being harmful. Broome, in his discussion of offsetting one's greenhouse gas emissions, suggests we explain its permissibility in this way. He says that offsetting is a way of doing "no harm" (Broome 2012, p. 85), so satisfies one's duty not to harm.[6] MacAskill also takes this approach. In contrast with attempting to offset one's adultery by buying indulgences, he says, "through effective carbon offsetting, you're preventing anyone being harmed by your emissions in the first place" (MacAskill 2015, p. 140).

This account of offsetting's permissibility relies on a couple things: (i) an appeal to a standard constraint against doing harm, and (ii) the observation that there's no such constraint against doing things that would have been harmful were it not for some further action one takes to prevent that harm. Though (i) is controversial, it's widely enough held and defended that relying on it does not seem like a serious cost. In support of (ii), we can cite various cases where one can permissibly do something that would be harmful were it not for some other action one has performed or would go on to perform.[7] Suppose I pick up a loaded gun, aim it at someone, and pull the trigger. Normally, pulling the trigger would be an act that is wrong because harmful. But in this case, it turns out, I had carefully unloaded the gun after picking it up, so the trigger pulling wasn't harmful. My unloading the gun prevented the trigger pulling from being harmful.

The components of the prevention account, then, are reasonably well supported. It also seems to be a good intuitive fit for the cases. After all, the end result of an offsetting case doesn't just leave the victim with the same amount of welfare they would have had, it leaves them either better off or else *exactly* as they would have been, irrelevant cambridge changes

---

[6]In his response to Cripps (2016), he makes this clear: "If you emit at one place, and also prevent an equal quantity of emissions at another place, *you do no harm* because because you do not change to the global concentration [of greenhouse gas]. This is how offsetting works." (Broome (2016, p. 159), emphasis added).

[7]Though certain of these cases may well cause trouble for maintaining (i); see Hanna (2015a,b).

aside. So how could one have done harm? One *would have* done harm, of course, had one not offset, but this just goes to show that what offsetting does is prevent something from being harmful.

Attractive as the prevention account of offsetting initially seems, it can't be right. The problem is that it would require that the offsetting be linked in the right way to the particular would-be harmful acts that the offsetter performs. I have to unload *my* gun to prevent myself from doing harm. If I unload someone else's gun, that might prevent someone else from doing harm—a good thing, to be sure—but it doesn't prevent my trigger-pullings from harming. So the prevention account requires that offsetting actions interfere in some way with the effects of the would-be harmful actions they are supposed to offset. But in many permissible offsetting cases, including our original cases, there is no such link to be found.

In ZAP OFFSETTING, the weight A removes is a *different* weight from the one they placed on the scale, one that was there before A came on the scene. Removing it doesn't affect at all what the weight A put on the scale does. It doesn't affect it any more than it affects what the other remaining weights do. The weight A added is still on the scale at the time of the zap, still making the zap worse than it would have been had the weight been absent. If A wanted to prevent their earlier action from harming, they should have made sure to remove the very weight they placed on the scale. But it seems there's no requirement to do this—removing one is just as good a way of offsetting as removing any other of equal or greater weight. Similarly for other cases. As Broome (2012, p. 85) himself notes, carbon offsetting "does not remove the very molecules that you emit," and that the molecules of CO_2 one has emitted "will wreak their damage" (p. 89), even when one has fully offset one's emissions. Offsetting actions don't need to affect in any way the consequences of the very actions they offset. So offsetting cannot be understood as preventing the harm that one's other actions would have done.

The prevention account has enough intuitive pull that it's worth dwelling on this point. Let's consider ZAP OFFSETTING in more detail. A places a 1 lb weight, call it w, on the scale. If A doesn't go on to remove some weight or otherwise offset, this has clearly harmed C—it makes the pain C suffers worse, just like in ZAP INCREASE.[8] And nothing about what the w does is changed by A's offsetting action of removing some other weight from the scale. There remains a rather direct causal path we can follow from the action to the worse pain. A puts w on the scale, w exerts downward force on the scale, which at the relevant moment contributes to higher voltage just like the other weights on it do, which makes for worse pain. Ordinarily, being able to trace such a causal path suffices for determining causation. And there's no preemption, overdetermination, or other factors which might undermine thinking w, like each of the other weights, is playing a causal role in the increasing the voltage: as with each of the other weights, if w were not there, the voltage would have been lower. The presence of w increases the harm to C, and it does so whether or not A has removed

---

[8]This is clearly harming on a counterfactual comparative account, since C would have been better off had A not placed the weight there. It's also true if we accept any plausible non-counterfactual comparative theory of harming. On a very flatfooted theory of harming as causing a harm, at least with reasonable assumptions about causation and the identity conditions of events, adding the weight doesn't harm, since there is no harmful event it causes, only one it contributes to. But this is just a problem for such a flatfooted theory. A more sophisticated version of the theory, such as the one Harman (2009) ends up with, will count the addition as a case of harming, since it causes C to "be in a particular bad state rather than a better state" (Harman 2009, p. 150).

What about accounts like Foot's (Foot (1967), Foot (1984)) or more recently Woollard's (Woollard (2008), Woollard (2015)), which classify actions as harmings depending on their role in some harmful sequence? This will depend on the details. Foot's theory is silent on these cases. Adding a single weight neither originates nor sustains, in the relevant sense, a harmful sequence, but nor does it seem to be a mere enabling or forbearance from prevention. One might easily extend the account, though, to include a classification of increasing or encouraging a harmful sequence as a case of harming. Woollard's account would treat the addition as a case of harming, since it is part of the sequence leading to the harm to C. But if I understand it correctly, it would also treat the removal of a weight as a case of harming for the same reason, which is unacceptable. The account could be modified, though, along the lines which Harman modifies the simple causal account by treating only worsenings of harms as harmings.

some other weight. So how could putting w on the scale not, in the end, harm C? I don't see a way of plausibly denying that it does. I conclude the prevention account fails to explain the permissibility of offsetting.[9]

3 Offsetting Not an Old Proviso

Even if I am right that offsetting cases involve doing harm, this does not yet show that permissible offsetting is incompatible with a constraint against doing harm. Practically nobody thinks that *every* possible harmful action is impermissible. If, somehow, the only way to prevent 1,000 innocent people from being killed is to punch some other innocent person, the punching is harmful yet permissible. It is also often permissible for surgeons to operate, even if they must do some harm in the process, so long as they expect the procedure to do enough good for the patient on the whole.

There are at least two options available for holding onto a constraint against harm while allowing that these harms are permissible.[10] One is to say that the constraint is not one simply against doing harm, but is rather against something more complicated, like doing harm that is unnecessary

---

[9]We can also make this point by observing that in certain cases of offsetting, the harm increasing action and everything in the causal chain between it and the harm to the victim can be intrinsically identical to that of a case with a harm increase that is not offset. Suppose that A removed the weight before adding their own (they are allowed to remove a weight without adding one, but if they don't go on to add one they must pay the large fee). The removal will make the scale have 6 rather than 7 lbs. Then the addition brings it from 6 to 7 lbs. This still seems like a permissible offsetting case. But compare the weight addition here with that of another case, where the scale started with 6 lbs to begin with. Here adding a weight and doing nothing after would be a harm. But this act and everything between it and the zap to C could be an intrinsic duplicate of what happened in the offsetting case. And it is plausible that if for two possible actions everything in the causal chain from the action through some harmful event is intrinsically identical, then one of these actions increases that harm iff the other does (cf. Paul and Hall (2013), especially Ch. 3 §4.3). So, since there is harm in the non-offsetting case, there is harm in the offsetting case. So offsetting can't be prevention of would-be harm.

[10]See, e.g., Kagan (1998a, §§3.2–3.3).

to prevent some much worse outcome.[11] Another option is to say that the constraint against harm is violated in cases like these, but that the reason it provides against doing the action can be *outweighed* by considerations of general goodness or benefit to the victim.[12] In either case, if a harmful act is necessary for bringing about some amount of goodness, either a large total amount or a potentially smaller amount to the victim of the harm, the act can thereby be permissible. And there are various other proposed provisos to the constraint against harm, such as that doing harm can be permissible if consented to, if deserved, and so on. Why not think offsetting cases can be subsumed under some well-known proviso?

Offsetting has features which make the plausible candidates unsuitable. We've seen from the cases we've considered that increasing harm and offsetting can be permissible even when the benefit from doing so, as opposed to not interfering at all, is negligible. This goes for benefits to the agent, to the victim, or in terms of impersonal total goodness. In ZAP OFFSETTING, there is no net benefit for C, a minor benefit for A, and no benefit to anyone else. So provisos which say you may harm if doing so is required for accomplishing some significant amount of good (or preventing some sufficient amount of bad)—either for oneself, for the victim, or in general—won't work. Compare: a doctor may permissibly amputate your leg without your permission in order to save your life, but assuming your legs are of roughly equal value to you, they cannot amputate your right leg in order to save your left without your permission, even if they would get some moderate benefit from doing so. It would be permissible to do this only if there were a rather large benefit to you or to others.

Note also that when a harm is to be justified by some benefit, it must be *required* for producing that benefit. Or rather, it must be required that there is some harm that is at least as bad. A doctor cannot permissibly amputate

---

[11]Or some subtler variant of this idea, like the Doctrine of Initial Justification from Kamm (2007, Ch. 5).

[12]Cf. the Tradeoff Idea from Thomson (1990, p. 123).

your leg simply for their own modest benefit even if they go on to save your life through some unrelated procedure. But in the offsetting cases we've considered the beneficial harm decrease does not require the harm increase, since *A* could have instead paid the fee to remove a weight.

The most promising provisos for making sense of offsetting, which concern a harm's relation to some benefit, won't save the constraint against harm. Other familiar provisos won't either. There is no consent from the victims in the cases we've considered. An appeal to hypothetical consent on its own doesn't help, since we'd need some further account of why offsetting would be consented to whereas, for example, having one leg amputated to save the other would not be. Moreover, it seems to me that the permissibility of increasing harm and fully offsetting it doesn't change even with explicit non-consent from the victims of the harm, in which case a hypothetical consent proviso can't help. And as we saw in §2, there need not be any causal connection between the harm decrease and the harm increase it offsets, so we can't treat offsetting cases as ones of withdrawing one's own aid.[13]

Perhaps there are other provisos which have been proposed which could explain how the permissibility of offsetting is compatible with a constraint against doing harm. But none that I am aware of can, so I think we should explore other options. I hope, at any rate, that the force of the puzzle I am interested in can now be felt. Offsetting cases involve doing harm and so violate the familiar deontological constraint against doing harm, and not in a way covered by standard provisos. So why are they permissible?

---

[13]For discussion of withdrawing aid, see McMahan (1993) and references therein, as well as Woollard and Howard-Snyder (2016, §7) for an overview of the more recent literature.

11

4 A Recipe for Offsetting

It will help to first think about what makes an offsetting case an offsetting case. Not all the features of our two cases are essential. Instead of removing a 1 lb weight, as in ZAP OFFSETTING, one could just as permissibly offset the weight addition by tying a −1 lb balloon to the scale. And there's nothing special about weights and zaps, of course. *A* acts permissibly, for example, in the following.

> **POISON OFFSETTING**
> *C* will get a headache from drinking out of a well which *B* has poisoned. How painful it will be depends on the amount of poison in the well. Currently there are 100mg of it. *A* adds 10mg of poison and an enzyme that will neutralize at least 10mg of the poison.[14]

So what features of offsetting cases are the essential ones?

Let's start with the obvious. One first needs some action that, at least on its own, would make some harm worse (adding a weight to the scale in ZAP OFFSETTING, adding poison to the well in POISON OFFSETTING). And one needs an action by the same agent that ensures that their total effect on the victim leaves them no worse off than they would have been had the agent not interfered at all (removing a weight, adding a balloon, adding the neutralizing enzyme).

As we can see from ZAP REPLACEMENT, however, the characterization thus far will not suffice.

> **ZAP REPLACEMENT**
> *A* prevents *B* from zapping *C*, but then goes on to zap *C* with the same voltage.

---

[14]Shouldn't *A* have just added the enzyme and not the poison? Suppose that *A* gets $1000 for adding poison, but must pay $950 for the enzyme .

Here A's total behavior leaves C just as well off as they would have been, but their zapping C seems impermissible. So we'll need a bit more complicated characterization to exclude this and similar cases from counting as offsetting.

I take the general structure of offsetting cases to be roughly this: independently of what some agent does, there's an impending intrinsically harmful event.[15] How bad it will be causally (not constitutively) depends on the magnitude of some quantity at some time: the more of the quantity at that time, the worse the harm. Before that time, the agent can alter that quantity both in ways that increase it and in ways that decrease it. Offsetting occurs when the agent does something to increase the magnitude of the quantity, but also does something that decreases it by at least as much, before the relevant time. The result is that the harm is not worse than it would have been had the agent not interfered.

This new characterization still applies to the offsetting cases. In Zap Offsetting, the intrinsically harmful event is the pain C will have from B's zap, which was going to occur independently of what A does.[16] The quantity is the weight on the scale, which A can increase by adding weights and decrease by removing weights. In Poison Offsetting, the intrinsically harmful event is the pain from the headache C will get from drinking the poisoned water and the quantity is the amount of un-neutralized poison in the well, which A can increase by adding more poison and decrease by adding the neutralizing enzyme. Having seen the pattern, it's easy to

---

[15]I'll be speaking of harms as events. This is controversial (Bennett (1995, Ch. 2), Hanser (2008), Thomson (2011), and Hanser (2011)), but not in ways that should matter here.

An event is intrinsically harmful to someone if that event constitutively contributes negatively to that person's welfare. This is meant to be analogous to the more familiar notion of intrinsic value (or disvalue) in the 'as an end' sense, rather than the 'dependent only on intrinsic properties' sense (see Korsgaard (1983), Kagan (1998b), among others).

I'll not theorize about 'impending' and 'independently' here, but instead leave them with their ordinary meanings. A more fully developed theory would require more precise characterizations.

[16]To say that *this* event would have occurred regardless of what A does is to assume that events aren't especially modally fragile (on which, see Lewis (1986, pp. 196–199)).

generate new offsetting cases: just plug in different kinds of impending harms, different kinds of determining quantities, and different ways for the agent to affect the quantity's magnitude.

In contrast with the offsetting cases, in ZAP REPLACEMENT, there is no impending harmful event independent of what A does. The harm that comes to C—the pain from A's zap—comes entirely from what A does. So our recipe does not classify this as a case of offsetting, as desired.

Even if the recipe is the right one, though, it still doesn't yet tell us why increasing harm is permissible when offset. Let us now take up that question.

5   A New Constraint

Here's my suggestion: drop the familiar constraint against doing harm and replace it with a constraint against unoffset harm increases. If we do this, we can explain how the harms in ZAP OFFSETTING and other offsetting cases can be permissible whereas the harm in ZAP INCREASE is not. The former will not violate this new constraint, whereas the latter does.

How would such a constraint look? One formulation just adds a conditional qualification to an ordinary moderate constraint against increasing harm:

> **Conditional Offsetting Constraint**
> Don't increase harm if you have not and will not fully offset it.

But it will need to specify what it is to offset a particular harm increase.

Looking to our recipe for offsetting cases as a guide, we might propose the following. Where α is an act performed by an agent X that increases the badness of harm h solely through increasing an intermediary quantity q by degree d, α is fully offset iff there is some act β that X performs that decreases h through decreasing q by at least d. So in ZAP OFFSETTING, A's harm increase

14

is fully offset because their removing another weight decreases the badness of the pain from the zap through removing as much weight from the scale as their harm increase added.

This proposal founders on slightly more complicated cases. Suppose in Zap Offsetting A had added *two* 1 lb weights but still only removes one 1 lb weight. We want our constraint to rule out such behavior. A hasn't sufficiently offset here, and would need to remove 2 lbs worth of weights to do so. But on the current proposal, each of A's acts would be allowed. For each weight addition A makes, there is an act—namely, the single 1 lb weight removal—that is sufficient for full offsetting according to the current proposal. The single harm decrease is being counted twice, offsetting two increases each as large as the decrease is.

Trying to avoid this by linking specific increases to specific decreases runs into trouble when the increases and decreases vary in size. A better strategy is to take an individual's decreases to 'distribute' across all of their increases. So if they place a 1 lb and 2 lb weight on the scale, a -1 lb balloon offsets their first weight by $\frac{1}{3}$ lbs and their second by $\frac{2}{3}$ lbs. To fully offset each harm, another 2 lbs worth of decrease is required.

Now that we are looking at total increases and decreases, though, an alternative formulation of the constraint seems simpler:

> **Holistic Offsetting Constraint**
> Don't allow your net contribution to a harm be positive.

On this proposal, it's the whole pattern of actions that can violate the constraint, rather than any particular components of that pattern. It need not take on commitments about which increases are or are not fully offset. When A adds two 1 lb weights but only removes 1 lb's worth, *Holistic Offsetting Constraint* doesn't say that either weight addition was a violation of the constraint. Instead, what violates the constraint is A's allowing their behavior as a whole to have the net effect that the harm is worse.

15

Though formulating the constraint this way takes us further from the familiar constraint against doing harm than *Conditional Offsetting Constraint* did, there's nothing too exotic here. It's akin to what we might say about failing to keep a promise. What's wrong, we might say, is neither the act of promising itself nor any of the acts one performs instead of fulfilling the promise, but rather the pattern of behavior as a whole.

Holistic Offsetting Constraint gets the permissible offsetting cases right. In neither Zap Offsetting nor Poison Offsetting does A make a positive net contribution to the harm. It also gets the partial offsetting cases right. However it's distributed across various actions, if A ends up adding more weight to the scale than they take away, A makes a positive net contribution to the harm, so runs afoul of the constraint. And it gets Zap Replacement right. When A zaps C, there is a harm—the pain they caused C—to which they have made a positive net contribution, so they violate the constraint. This is so even if the net contribution of their behavior to C's overall welfare is neutral or positive. What the constraint unoffset harm increases cares about is not one's total contribution to a particular person's welfare, but rather one's total contribution to a particular harm.

Both the holistic formulation and conditional formulation with proportional distributive linking seem to work for all the offsetting cases we've considered. There is, though, a potential problem remaining. A harm can be worse because it involves pain that is more intense, or because it lasts longer, or because it is more widespread; intensity, duration, and (apparent) spatial extent are all what I will call *aspects* of the harm that comes from pain. This means that a harm could be increased in one aspect and decreased in another. But just as one can't permissibly offset increases to one harm by decreasing another, as in Zap Replacement, it seems to me that harm decreases can't offset harm increases to the different aspects of the same harm. If the pain of the zap were determined by two scales, one which determined the voltage of the zap and the other determined its length, I

don't think it would be permissible for *A* to move weights from one to another without *C*'s consent, making the pain longer but less intense, even if the result is no worse than it would have been otherwise.

The current formulations of the constraint, however, don't make any distinction sensitive to this, and so would seem not to be violated by this kind of inter-aspect tradeoff. Thus I suggest we revise the constraint to be either

> *Conditional Offsetting Constraint (Aspectual)*
> Don't increase any aspect of harm if you have not and will not fully offset it,

or

> *Holistic Offsetting Constraint (Aspectual)*
> Don't allow your net contribution to any aspect of a harm be positive.

Though a full theory including such a constraint would require an account of what exactly an aspect of a harm is,[17] not to mention an account of what exactly a harm is, I think we can already see that it will sort the relevant cases correctly, and so is a promising way to explain the permissibility of offsetting.

6   Deriving What the Old Constraint Gets Right

The old constraint against harm is popular for a reason: there are plenty of ordinary cases of harming that it correctly rules out as impermissible. If we

---

[17]Here is a first pass: an aspect of a harm is a trope the existence of which is an ultimate normative partial ground of the fact that the harm is as intrinsically harmful as it is. But even granting this talk of ultimate normative grounding (see Fine (2012) and Bader (2017), but also Berker (2017)), this only gets us close to what we're after. It makes a pain event's being a half-second long an aspect of the harm, whereas what we want is that the event's *temporal length* to be an aspect. I leave patching this up for elsewhere.

are going to replace the old constraint with a new one, we'd better make sure that we can still predict the wrongness of ordinary harms.

Happily, we don't need to make any modifications to the new constraint to do this, since it is violated by ordinary harmful actions. If I punch somebody, I make a net positive contribution to all of the aspects of the harm that comes to them from being punched. After all, I contribute all of the harm, and there's certainly a positive amount there. Could fully offsetting it make my harm increase in such a case permissible? Well, no. But this is no problem for the proposed derivation, since the reason offsetting can't make it permissible isn't because offsetting fails to do the normative work it does in the permissible offsetting cases, but rather because it's impossible in this case to fully offset the harm increase but still have done harm. If the increase to harm had been fully offset, there would have been no harm at all.

A harm of the familiar kind, then, is also a harm increase that is not fully offset. Thus—putting aside cases where offsetting happens—any violation of the old constraint against harm will also be a violation of the new constraint against unoffset harm increases, so we are not losing what the old constraint got right by moving to the new constraint. We should replace the traditional constraint against harm with a constraint against unoffset harm increases.

Sometimes the problem with a moral theory is that it misses a morally relevant distinction entirely, sometimes that it gives weight to a distinction not worth caring about. But other times the problem is more subtle: the theory has drawn a distinction which is *close* to a morally important one, but isn't quite carving at the moral joints. Theories that make this kind of error can be very plausible, since they may classify plenty of cases correctly, even if not for exactly the right reasons. But chances are they'll go wrong somewhere, potentially in very significant ways.

I've argued that theories which incorporate the usual kind of constraint

against doing harm are making this kind of error. In correctly ruling out most cases of harm as impermissible, they overshoot, also ruling out cases of fully offset harm increases, like *A*'s action in ZAP OFFSETTING. A theory with a constraint against unoffset harm increases does better.

The remainder of the paper briefly explores two kinds of implications of a shift to the kind of constraint I prefer.

7 The Source of the Constraint

Ultimately we want an account not just of what deontic constraints there are, but also an account of why there are the constraints that there are. Moving from a constraint against harm to one against unoffset harm increases may require us to rethink where this constraint is coming from.

An attractive idea to many deontologists is that autonomy is the source of a constraint against doing (as opposed to allowing) harm. For a person to have genuine free control over their own mind and body, they should be protected against certain kinds of impositions. On the one hand, they should be protected against being causally imposed upon by others in certain ways. This is why there is a constraint against doing harm.[18] On the other hand, a person also needs to be free from certain kinds of normative impositions, like being morally required to intervene on another's behalf. This is why there is no constraint against allowing harms.[19] I am cutting a long and controversial story short here, since my aim isn't to work through any detailed version of this view, but just to raise the question of whether something like it can be used to derive a constraint against unoffset harm increases.

It may seem easy enough for an autonomy theorist to explain why

---

[18]See Quinn (1989), Shiffrin (2012), and Woollard (2015, Ch. 6).

[19]For discussion, see Slote (1985, pp. 23–34), Kagan (1989, pp. 236–241), Shiffrin (1991), and Woollard (2015, pp. 107–111)

increasing harm and offsetting it would be permissible. *A*'s behavior as a whole in Zap Offsetting leaves *C* exactly as well off as they would have been had *A* not done anything to interfere. And it's not just that there's no overall difference in welfare for *C*, it's that there's no negative difference at all in the harm *C* suffers, but only irrelevant cambridge changes or, in cases like Zap Offsetting (Negative), improvements. And since *A*'s not interfering at all would not impose upon *C*, it's plausible that *C* has not been imposed upon by *A* when *A* increases but fully offsets the harm, either.[20] And since there is no such imposition, it would be an unjustified normative imposition on *A* to require them not to behave in this way. So what *A* did in the offsetting case should be permissible.

But we are not done yet. We need to explain not only why doing harm in offsetting cases is permissible, but also why doing harm is impermissible in others, including cases where the harm one does replaces another harm or lack of benefit of the same size, as in Zap Replacement or cases of harming followed or preceded by compensation. This is more difficult, since it seems we can make a very similar argument for the permissibility of harm replacement. What *A* does in the replacement cases leaves *C* just as well off as they would have been had *A* done nothing. So how has *C* been imposed upon here any more than in the offsetting cases?

One could say that while all of *A*'s actions together don't make *C* worse off, one of them (the zap) does, which is enough of an imposition to justify making it impermissible. But this won't do, since the same point could be made about the offsetting case if we single out the harm increasing act.

We might try appealing to the point that in offsetting cases, it's not just "no difference in the victim's welfare", but "no difference at all in the harm *C* suffers, improvements and irrelevant cambridge changes aside". Even if you should be indifferent between not being zapped and being

---

[20]Note that we need to be considering the agent's behavior as a whole here, rather than its proper parts (cf. Portmore (2017) and Brown (2018)).

zapped but given \$200, it's not implausible to say that it's a violation of your autonomy if I exchange one for the other without your permission. But what about Zap Replacement? We can make the harms in question intrinsically identical, not just equal in badness, so how can replacing one with the other be a morally relevant imposition?

It seems we need some way of explaining why the identities of harms would important: why replacing this harm with that qualitatively identical one would be impermissible, whereas adding and subtracting the same amount from it can be permissible. It is unclear how an appeal autonomy can make sense of this.

This is not to say that we should give up on autonomy as a source of the constraint against unoffset harm increases. I think it remains a promising direction to pursue. But we may find that some alternative theory does better. In any case, more work is required to derive the constraint against unoffset harm increases from deeper principles, and current frameworks may require revision or replacement in order to do so. Moving to such a constraint may thus have theoretical implications that go beyond questions about how best to formulate a first-order deontic theory.

8   CO$_2$ Emissions and Meat Purchases

Moving to a constraint against unoffset harm increases may also make a difference to what is permissible in practically important, real-life cases.

In the climate change case, it's plausible that the harms from emitting CO$_2$ are ones which come solely from increase to the global concentration in CO$_2$, as Broome (2012, 2016) claims. And the global concentration of CO$_2$ is a quantity that one can increase or decrease. One could worsen the harms of climate change by emitting CO$_2$, but also reduce those very harms by offsetting those emissions. So it appears to fit our recipe for offsetting cases, opening up the possibility of permissible offsetting, though there

will be tricky issues about timing and the identity of harms we would need to consider. The question will be whether by emitting and buying carbon offsets one's behavior will involve a net increase to any aspect of any harm.

In the animal product case, it is plausible that the most significant harms are the ones that come solely from increased demand for these products from specific suppliers within specific windows of time. Plausibly, it is only the amount of demand for these products as perceived by the relevant suppliers at the relevant times which determines how bad the harms are. One can increase this demand through purchasing, but perhaps there are things one could do—paying others not to purchase, for example—to decrease this demand.[21] So it seems plausible that this might fit the offsetting recipe, again opening up the possibility of permissible offsetting.

It would take a good deal of empirical and philosophical work to sort out the details of these cases. We cannot hope to settle here whether they could involve permissible offsetting and if so, how exactly that offsetting can be done. We have laid some important groundwork for approaching these questions, however. If what I've argued is right, what we need to look for is not whether any potential means of offsetting will prevent our actions from harming or help them meet any of the usual provisos to a constraint against harm. Rather, we must look to see whether our behavior as a whole will have the net effect of worsening a harm in any way. If it does, we act (pro tanto) wrongly. It may be, though, that we can act in ways that worsen harms but act in other ways which offset those worsenings. In this case our response, other duties aside, need not be to refrain from these harmful actions, but rather to ensure they are fully offset.[22]

---

[21]This would be difficult to implement. One would need to know that someone was going to make a particular kind of purchase, that paying them would keep them from making that purchase, that they will not just buy something else objectionable instead, and that they weren't disposed to make the purchase in the first place because they expected you to pay them not to. Similar complications go for preventative offsetting of carbon emissions.

[22]One may have duties from other sources to do more than this. Besides offsetting one's

References

Bader, Ralf (2017). "The Grounding Argument Against Non-reductive Moral Realism". In: *Oxford Studies in Metaethics*. Ed. by Russ Shafer-Landau. Vol. 12.

Bennett, Jonathan (1995). *The Act Itself*. Oxford: Clarendon Press.

Berker, Selim (2017). "The Unity of Grounding". In: *Mind* 127, pp. 729–777.

Broome, John (2012). *Climate Matters: Ethics in a Warming World*. W. W. Norton & Company, Inc.

— (2016). "A Reply To My Critics". In: *Midwest Studies In Philosophy* 40, pp. 158–171.

— (2019). "Against Denialism". In: *The Monist* 102, pp. 110–129.

Brown, Campbell (2018). "Maximalism and the Structure of Acts". In: *Noûs* 52.4, pp. 752–771.

Cripps, Elizabeth (2016). "On *Climate Matters*: Offsetting, Population, and Justice". In: *Midwest Studies In Philosophy* 40, pp. 114–128.

Fine, Kit (2012). "Guide to ground". In: *Metaphysical Grounding: Understanding the Structure of Reality*. Ed. by Fabrice Correia and Benjamin Schnieder. Cambridge University Press, pp. 37–80.

Foot, Philippa (1967). "The Problem of Abortion and the Doctrine of Double Effect". In: *Oxford Review* 5, pp. 5–15. Reprinted in Foot (1977/2002, pp. 19–32).

— (1984). "Killing and Letting Die". In: *Abortion and Legal Perspectives*. Ed. by Jay L. Garfield and Patricia Hennessey. University of Masachusetts Press. Reprinted in Foot (2002, pp. 79–88).

— (2002). *Moral Dilemmas: and Other Topics in Moral Philosophy*. Oxford University Press.

---

own emissions, one may be morally required to push for international political solutions. Besides offsetting animal product consumption, one may be required to promote laws requiring humane treatment or advocate that others offset their consumption as well. And there may be additional duties to not participate at all, such as a duty to avoid complicity through benefiting from wrongdoing (McPherson 2018).

Foot, Philippa (1977/2002). *Virtues and Vices and Other Essays in Moral Philosophy*. Oxford University Press.

Hanna, Jason (2015a). "Doing, Allowing, and the Moral Relevance of the Past". In: *Journal of Moral Philosophy* 12, pp. 677–698.

— (2015b). "Enabling Harm, Doing Harm, and Undoing One's Own Behavior". In: *Ethics* 126, pp. 68–90.

Hanser, Matthew (2008). "The Metaphysics of Harm". In: *Philosophy and Phenomenological Research* 77, pp. 421–450.

— (2011). "Still More on the Metaphysics of Harm". In: *Philosophy and Phenomenological Research* 82, pp. 459–469.

Harman, Elizabeth (2009). "Harming as Causing Harm". In: *Harming Future Persons*. Ed. by Melinda A. Roberts and David T. Wasserman. Springer, pp. 137–154.

Herzog, Howard J. (2018). *Carbon Capture*. The MIT Press.

Kagan, Shelly (1989). *The Limits of Morality*. Oxford: Clarendon Press.

— (1998a). *Normative Ethics*. Westview Press.

— (1998b). "Rethinking Intrinsic Value". In: *The Journal of Ethics* 2.4, pp. 277–297.

Kamm, F. M. (2007). *Intricate Ethics: Rights, Responsibilities, and Permissible Harms*. Oxford University Press.

Korsgaard, Christine M. (1983). "Two Distinctions in Goodness". In: *The Philosophical Review* 92.2, pp. 169–195.

Lewis, David (1986). "Events". In: *Philosophical Papers*. Vol. II, pp. 241–269.

MacAskill, William (2015). *Doing Good Better*. Penguin Random House.

McMahan, Jeff (1993). "Killing, Letting Die, and Withdrawing Aid". In: *Ethics* 103, pp. 250–279.

McPherson, Tristram (2018). "The Ethical Basis for Veganism". In: *The Oxford Handbook of Food Ethics*. Ed. by Anne Barnhill, Mark Budolfson, and Tyler Doggett, pp. 210–240.

Nefsky, Julia (2019). "Collective harm and the inefficacy problem". In: *Philosophy Compass* 14 (4).

Paul, L. A. and Ned Hall (2013). *Causation: A User's Guide*. Oxford University Press.

Portmore, Douglas W. (2017). "Maximalism Versus Omnism about Permissibility". In: *Pacific Philosophical Quarterly* 98.S1, pp. 427–452.

Quinn, Warren S. (1989). "Actions, Intentions, and Consequences: The Doctrine of Doing and Allowing". In: *The Philosophical Review* 98.3, pp. 287–312.

Shiffrin, Seana Valentine (1991). "Moral Autonomy and Agent-Centered Options". In: *Analysis* 51.4, pp. 244–254.

— (2012). "Harm and its Moral Significance". In: *Legal Theory*, pp. 1–42.

Slote, Michael (1985). *Common-sense Morality and Consequentialism*. Routledge & Kegan Paul.

Thomson, Judith Jarvis (1990). *The Realm of Rights*. Harvard University Press.

— (2011). "More on the Metaphysics of Harm". In: *Philosophy and Phenomenological Research* 82, pp. 436–458.

Woollard, Fiona (2008). "Doing and Allowing, Threats and Sequences". In: *Pacific Philosophical Quarterly* 89, pp. 261–277.

— (2015). *Doing and Allowing Harm*. Oxford University Press.

Woollard, Fiona and Frances Howard-Snyder (2016). "Doing vs. Allowing Harm". In: *The Stanford Encyclopedia of Philosophy*. Ed. by Edward N. Zalta. Summer 2016. http://plato.stanford.edu/archives/sum2016/entries/doing-allowing/.