Offsetting Harm[*]

[feel free to cite, but beware: final version may differ]

Mike Deigan
mike.deigan@rutgers.edu

All of us act in ways that increase harms. By driving, buying plane tickets, and even by exhaling, I increase the amount of CO_2 in the atmosphere. 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. Aside from polluting, we have many other opportunities to contribute to processes which result in more or less harm, depending on how they are contributed to. We can buy animal products produced by factory farms, increasing demand for them, we can join in social or physical pile-ons, worsening their effects, and so on. Given their ubiquity and outsize effects on others, it is practically important to know how we can permissibly interact with such processes.

Assuming we're talking about cases where one's actions do in fact perceptibly increase some harm, we might think that there is no special theoretical difficulty here. 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 simply apply this theory to the cases of increasing harm that are practically significant. But I think this is a mistake. 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

[*] Thanks to Stephen Darwall, Keith DeRose, Daniel Ferguson, Moya Mapps, Juan Piñeros Glasscock, Daniel Moerner, Zoltán Gendler Szabó, Victor Tadros, Yuan Yuan, two anonymous reviewers the members of the Fall 2016 Yale Philosophy WIP seminar, and especially to Shelly Kagan for very helpful discussions of the ideas of this paper.

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, we can sometimes also contribute to them in ways that *decrease* those harms. We can buy carbon offsets, promote veganism or pay others not to buy animal products, we can mitigate the degree of social or physical harms, and so on. This means that sometimes we can act in ways that increases a harm, but still have total contributions *to that very harm* that are net neutral or negative. We can *offset* some harm increases. And often, it seems permissible to increase a harm so long as one offsets that harm increase.

This paper is an exploration, from the perspective of more-or-less commonsense deontological theory, of why offsetting harm is permissible.[1] I will argue that the standard deontological constraint against doing harm cannot accommodate permissible offsetting, and so should be replaced.

1 Zap Offsetting Cases

To see the puzzle about permissible offsetting, let's think about some artificially simple cases.

The innocent *C* will soon be zapped by 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

[1] The 'commonsense deontological theory' I'll be assuming holds that while there is a pro tanto duty to promote the general good, there are moderate constraints that sometimes require agents to not do so, as well as options both of the agent-sacrificing and the agent-promoting kind. These assumptions are very controversial, of course, but nevertheless commonly enough held for a project which takes them as starting points to be of interest. And they should be of interest even for those with doubts about such assumptions (indeed, I count myself among this group). The best way to find out which sort of ethical framework is right, it seems to me, is to work out in a fair amount of detail the best theories within these frameworks, and compare them with each other. I see the project which this paper begins as a small part of attempting to work out the best version of a deontological theory.

zap. One can remove a weight from the scale, *but only if one adds a weight of one's own*. And there are no other ways to interfere with the zapping. (The mechanism for ensuring this is left to the reader's imagination.) 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 three variants.

> **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* knowingly makes the harm to *C* worse without good reason, and in so doing acts impermissibly.

> **Zap Offsetting (Negative)**
> *A*—a collector of decorative weights—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.

In contrast, *A*'s action here is clearly permissible. 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. But is net harm reduction necessary for offsetting the weight addition in a way that makes it permissible? I think not. Neutrality seems to suffice, as can be seen in the apparent permissibility of what *A* does in Zap Offsetting.

> **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, and they are blue. *A* places the red weight on the scale and removes a blue one, leaving the total weight on the scale at 7 lbs.

It seems, then, that one can do something that makes a harm worse (adding a weight to the scale), so long as one offsets it by reducing the harm by at least as much. The question I'm interested in is why offsetting of this kind is permissible, given that *A*'s action in Zap Increase is impermissible.

We might wonder whether there is really any special puzzle here worthy of investigation. Isn't there a very simple answer 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* for no good reason, so acts impermissibly. But in the offsetting cases, *A* doesn't do any harm (or violate any other constraints), so acts permissibly. Simple as that, right?

Wrong. I will argue that this straightforward account fails, since one *does* do harm in offsetting cases. 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. On my view, one may contribute to a harm in a positive way—in a way that worsens it—so long as one offsets this contribution by doing something that leaves one's *net* effect on that harm neutral or negative. Doing something that increases a harm is not itself 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.

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 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 as one emits?[2] And given that

[2] How one can do this is more complicated than planting trees, or paying others to do so, since as Broome notes, once trees die they typically decompose and return the carbon they've stored back into the atmosphere. There are ways to artificially extract and

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?[3] Whether such behavior is permissible, I think, will depend on what makes offsetting permissible, and whether these cases have the relevant features. It is practically important, then, to figure out what makes offsetting permissible.

Here's the plan for the rest of the paper. I'll begin by laying out some general features of offsetting and considering the proposal, suggested by Broome (2012), that the reason the 'harm' one does in offsetting cases is permissible is because it is no harm at all, since what offsetting does is prevent one's other actions from doing harm. I argue that such a proposal cannot work. I then develop my own account, which replaces the constraint against doing harm with one against increasing harm without offsetting it. I show that this constraint gets the offsetting cases right and can also account for the wrongness of more familiar kinds of harm, then briefly consider what implications it has for carbon and meat offsetting. I conclude by considering a couple of objections and putting forward some ideas about what deeper explanation there might be for the permissibility of offsetting.

2 A Recipe for Offsetting Cases

What makes an offsetting case an offsetting case? Not all the features of our first 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

store carbon in more or less permanent ways, but these are currently rather impractical, at least on a large scale (see Herzog (2018, Ch. 6)). Broome thus 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).

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

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, but then adds an enzyme to the well that will neutralize 10mg of the poison.[4]

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.

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

[4] Shouldn't *A* have just added the enzyme and not the poison? Suppose that *A* gets a reward for adding poison, but must pay for the enzyme.

event.[5] 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.

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.[6] The quantity is the weight on the scale, which A can increase by adding

[5] I'll be speaking of harms as events. This strikes me as fairly natural, but is controversial. See, e.g., Bennett (1995, Ch. 2), Hanser (2008), Thomson (2011), and Hanser (2011). Some details would go differently, and for some issues these will be of great importance, but for current purposes I don't think it will matter much if we carry out the discussion in terms of events instead of facts or states or tropes.

An event is intrinsically harmful to someone if that event, in itself or constitutively, contributes negatively to that person's welfare (though we may want to restrict this if we think not all bads are harms). 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. I agree with Korsgaard (1983) and Kagan (1998b), among others, in thinking these senses should be sharply distinguished, even if, as now seems doubtful, they turn out to in fact be extensionally, or even intensionally, equivalent.

I'll not theorize about 'impending' and 'independently' here, but instead leave them with their ordinary meanings. No doubt a fully developed theory would require more precise characterizations, perhaps with substantial deviations from the commonsense notions.

[6] To say that *this* event would have occurred regardless of what A does is to assume that events aren't especially modally fragile. Some philosophers would say that it wouldn't be the same event if it were any more or less painful, in which case A could have prevented it from occurring by making the machine zap with a tiny bit more or less voltage. Some might go farther and say that it wouldn't have been the same event if there's any difference in its causal history, in which case any interference by A would make it a different event. I take this view to be false (for an argument against very fragile events, see Lewis (1986, pp. 196–199)), but we could carry out the same discussion in different terms without assuming that it is. We would just need to introduce a counterpart relation that relates the events that I would take to be the same, and say that what would happen regardless of what A does is the intrinsically harmful event or some counterpart of it. We needn't quibble, at least at this point, over whether these events are *really* the same.

weights to the scale and decrease by removing weights from it. 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 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 seems desirable.[7]

We now have at least a rough recipe for permissible offsetting cases. But even if it's the right one, it still doesn't yet tell us why offsetting should be permissible. Let us now take up that question.

3 Offsetting as Prevention?

Before offering my own account of why the offsetting is permissible, I'll consider a natural proposal that I think we should reject. The proposal says offsetting cases are permissible because they don't really involve doing harm; what offsetting does is prevent an 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.[8]

[7] We will return to cases like this and their relation to offsetting in §8.

[8] 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). MacAskill also takes this

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.[9] 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. Or consider scenarios where we lead someone to rely on us. If for one of those trust-building exercises I promise to someone that I'll catch them when they fall backwards, and they let themselves fall because of this, it's plausible to say my promising harmed them if I stand by and don't catch them. Catching them prevents me from harming them, and so makes permissible what would have otherwise been impermissible (namely, my

approach to explaining carbon offsetting's permissibility:

> In buying indulgences, you don't "undo" the harm you've caused others or the sins you've done. In contrast, through effective carbon offsetting, you're preventing anyone from being harmed by your emissions in the first place: if you emit carbon dioxide throughout your life but effectively offset it at the same time, overall your life contributes nothing to climate change. Similarly, "offsetting" your adultery (even if you genuinely could) would still affect *who* is harmed, even if it keeps the total number of adulterous acts constant. In contrast, carbon offsetting prevents anyone from ever being harmed by your emissions; its the "equivalent" of never committing adultery in the first place.
>
> MacAskill (2015, p. 140)

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

promising).

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 aside. Given this is 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 and make sure *my* promises are kept in order to prevent myself from doing harm. If I unload someone else's gun or ensure someone else's promises are kept, that might prevent someone else from doing harm—a good thing, to be sure—but it doesn't prevent my trigger-pullings and promisings from harming. So the prevention account requires that offsetting actions interfere in some way with the effects of the 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 the other cases. In Poison Offsetting, there's no guarantee that the enzyme will dissolve the poison A added—indeed, it's overwhelmingly likely that some of what it will dissolve is the poison that was there already, and so will leave unaffected at least some of A's addition. And 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.[10] And nothing about what the w does is

[10] 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. Harman's trick is to treat harms as being extremely modally fragile, so that causing any difference causes a different harm. Then to prevent overgeneration (to keep, say, simply removing a couple weights from counting as doing harm), she restricts harming to cases of worsenings of harms, or of causing someone 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 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.[11] It seems, then, that the presence of w increases the harm to C. And it does so whether or not A has removed 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.[12]

easily modified, though, along the lines which Harman modifies the simple causing harm account, by treating only worsenings of harms as harmings.

[11] This qualification is an important one, as [redacted] has helped make clear to me. There are cases where it seems that we can trace the same kind of causal path from each weight to the zap, but where we intuitively think some but not others cause an increase. For example, suppose the weight increases the voltage until it reaches some maximum weight, at which point more weight makes no difference. Suppose this limit is 5 lbs. Then if A adds a weight making the total weight on the scale 8 rather than 7 lbs, it is natural to say A has not increased the harm, whereas some of the weights added earlier did. This despite the fact that the weight A added seems to be doing exactly what the other weights are doing. This case, it seems to me, is in some ways like cases of 'pure overdetermination' and in other ways like 'late preemption' (on which, see Paul and Hall (2013, pp. 99–160)). In particular, the relevant counterfactual seems to fail: it's not true in this case that if the weight A added weren't there, the harm would have been less bad.

These and similar cases are important for metaphysicians to consider, and perhaps what they find may justify overturning our judgements about cases which don't involve anything like overdetermination. In the meantime, however, I take it to be reasonable to rely in on our ordinary ways of determining causation in cases where we don't have reason to think there is overdetermination.

[12] We can also make this point by observing that in certain cases of offsetting, the harm increasing action and everything causally 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, for

Now the force of our puzzle can be felt: offsetting cases involve doing harm, and so violate the familiar deontological constraint against doing harm, so why are they permissible?

4 A New Constraint

If we want to account for the permissibility of offsetting, we'll have to look beyond the prevention account. 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: there is no constraint against harming, and the harming done in these cases doesn't violate the constraint against unoffset harm increases.

How would such a constraint look? As we'll see, there are some complications. But let's start with a simple formulation, which just adds a conditional qualification to an ordinary constraint against increasing harm:

> ***Conditional Offsetting Constraint***
> Don't increase harm if you have not and will not fully offset it.[13]

instance, that *A* removed the weight before adding their own. 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. 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. (I take this to be plausible enough on its own, but it could also be derived from the principle that causal effects supervene on intrinsic features of the causal pathway, for an elaboration and qualified defense of which, see 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.

[13] As with any deontological constraint, we will want this to be a moderate and not an absolute one. It's not that any conceivable unoffset harm increase is impermissible—suppose a small harm increase will save many lives, for example. We can make the

Increasing harm, I take it, we have enough of an intuitive grip on. It's just making some harm worse or introducing harm where there was none before.[14] But we need to say more about what it takes to fully offset some 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 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. Clearly, 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 using *Conditional Offsetting Constraint* and the current proposal about how offsetting works, it seems each of A's acts in this case 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. So each harm increase of A's would be permitted by *Conditional Offsetting Constraint*. The single harm decrease is being counted twice, offsetting two increases each as large as the decrease is.

To avoid this kind of double counting, it seems we need some way of linking the decreases with specific increases of an appropriate size. But how are we to do this? As we saw in considering the prevention account,

constraint moderate either by allowing it to permissibly violated in certain conditions, or else build the exception clauses into the constraint itself.

[14] We have enough of a grip on this for current purposes, but of course this will bring along with it the old debate about whether we can successfully draw a morally relevant doing/allowing distinction.

there need not be any causal link between them, so we can't look there for the requisite linking. We might try saying that increases and decreases are linked by temporal relations. Perhaps it's always the earliest harm increase that gets offset by the earliest equivalent harm decrease (which then cannot count towards offsetting other harm increases), so it's the second weight addition was the one that was impermissible. But this is arbitrary. We could just as easily say that it's the most recent unoffset harm that gets offset, in which case the first weight addition would be the impermissible one. It would be surprising if there were a brute fact which decides this one way or the other. And we get into more serious trouble once we start to consider separate but simultaneous acts. If *A* adds two 1 lb weights at the same time, but only removes 1 lb worth of weight, the temporal order linking principles won't work. Perhaps it's the earliest unoffset harm increase that's farthest to the agent's left that gets offset? No sane moral theory will include principles like this.

An alternative way of understanding how offsetting works can make the statement of the constraint we've been considering more palatable. Instead of requiring each decrease to be linked to some particular increase (or subset of increases), we say that the effects of a decrease of some harm are distributed evenly over all the agent's increases to that harm. So if *A* adds two 1 lb weights but only removes 1 lb worth of weight, neither of the harm increases are fully offset, so both violate *Conditional Offsetting Constraint*, though the amount of unoffset harm increase they each involve will be less than if *A* hadn't removed any weight.[15]

This idea about how to link harm decreases to increases is still underspecified, though, since there are different ways of spelling out what it is to

[15] Should we worry that we get *two* violations of the constraint here, whereas seemingly equivalent cases—like where *A* adds a 1 lb weight but doesn't remove any—only involve one? I think not. We should link the stringency of the constraint with the degree to which a given harm increase is not offset: the more the unoffset harm increase, the worse the wrong. We'd want to connect degree of harm with stringency anyways, so there's nothing too ad hoc about this (cf. the Aggravation Principle from Thomson (1990, p. 154)).

distribute something evenly. It could be that it subtracts an even *amount* of harm increase, or it could be that it subtracts an even *proportion*. Suppose A added a 2 lb weight and then a 1 lb weight, and removed 2 lbs worth of weights. Is an even distribution of the decrease to take away 1 lb's worth from each of the harm increases, or is it to take away $\frac{2}{3}$ lb's worth from the 2 lb addition and $\frac{1}{3}$ lb from the 1 lb addition? Note that these accounts would differ on whether the second addition violates the constraint—the amount version says it doesn't, the proportion version says it does.

I take the proportional construal to be preferable, partly because it seems less likely to run into the kinds of arbitrariness worries seen previously and partly because it doesn't require any revision to deal with cases where there is one large decrease and several smaller increases. For example, suppose A adds a 1 lb weight and a 2 lb weight and removes a 3 lb weight. The proportional construal correctly implies that both harm increases are fully offset and the constraint isn't violated. But the amount construal, as it stands, seems to say that this involves a partially unoffset harm, since it would distribute $1\frac{1}{2}$ lbs worth of harm reduction to each weight addition, leaving a $\frac{1}{2}$ lb's worth of harm remainder for the 2 lb weight addition. Perhaps we could add in some fix which allows remainders of offsetting decreases to be distributed over remaining harm increases, so this isn't a knockdown argument against the amount construal, but it does show it will need to be more complicated. With the distributive (as opposed to direct linking) proportional (as opposed to amount) understanding of offsetting, though, *Conditional Offsetting Constraint* seems get the cases right, and so is an attractive formulation of a constraint against unoffset harm increases.

We can avoid much of the above complication, however, by formulating the constraint in a different way:

> ***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's wrong, rather than any particular components of that pattern. It applies to the whole contribution that one makes to a harm, rather than to individual harm increases, so won't have to 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. It need not say that some particular act or omission is what constituted the violation.

Formulating the constraint this way takes it somewhat further from the familiar kind of constraint against doing harm than *Conditional Offsetting Constraint* is, but the idea is not an unfamiliar one. It is 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. Or consider failing to meet some moderate duty to aid: it's not that some particular action of yours constitutes the failure, it's that in all of your actions together, the required aiding action is not there. I don't think we should be bothered by putting our constraint into this kind of format.

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 various 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.

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 for either way of articulating the constraint that suggests they will need revision, or at least elaboration. 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 (like 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, 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, since they would not involve a positive net contribution to the harm as a whole. 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.

These constraints say we need to make sure not just that our behavior as a whole doesn't worsen any harm, but also it doesn't worsen any aspect of any harm. A full theory including such a constraint would require an account of what exactly an aspect of a harm is (not to mention an account of what exactly a harm is!), and I don't yet have such an account.[16]

[16] Here is a first pass: an aspect of a harm h is a trope the existence of which is an

Nevertheless, I think we have a good enough understanding of them to see that positing such a constraint is a promising way to capture our intuitions about offsetting cases.

5 Deriving What the Old Constraint Gets Right

So far I've argued that the permissibility of offsetting cases is not compatible with a constraint against doing harm, even taking prevention into account. And I've now given a couple ways of stating a new constraint against unoffset harm increases that does a better job. But the old constraint against harm is popular for a reason: there are plenty of ordinary cases of harming—like punching someone without good reason—that it correctly rules out as impermissible. If we 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 increase to harm 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 your increase to harm but still have

ultimate normative partial ground of the fact that *h* is as intrinsically harmful as it is. But even granting this talk of ultimate normative partial grounds makes sense (see Fine (2012) and Bader (2017) for the claim that there is a normative grounding relation distinct from other kinds of grounding and Berker (2017) for an argument against it) 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 really want is that the event's *temporal length* to be an aspect. I leave the question of how to patch up this account for elsewhere.

done harm. If the increase to harm had been fully offset, there would have been no harm at all.[17]

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 thus replace the traditional constraint against harm with a constraint against unoffset harm increases.

6 Applications: CO_2 emissions and meat purchases

Suppose I'm right that in order to make sense of offsetting cases like Zap and Poison Offsetting, we should reject the constraint against doing harm and instead accept one of the constraints against unoffset harm increases introduced above. What does this mean for practically important real-life cases? Would my constraints allow for offsetting of CO_2 emissions? Would they allow for offsetting of the purchase of animal products?

In the climate change case, it's plausible that, as Broome (2012, 2016) claims, the harms from emitting CO_2 are ones which come solely from increase to the global concentration in CO_2. 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. The question

[17] In fact, this would turn out not to be a case of offsetting, exactly, but rather a case of preventing one's own harm increase, similar to the proposal discussed in §3. Note that the problems for that proposal cannot be replicated for the present one, since those problems require that there be, in the end, some harm that comes to the victim.

will be whether by emitting and buying carbon offsets one's behavior will involve a net increase to any aspect of any harm, not whether we can successfully prevent ourselves from doing any harm at all.

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 for the relevant supplier at the relevant time which determines how bad the harm is. 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.[18] So it seems plausible that this might fit the offsetting recipe, again opening up the possibility of permissible offsetting.[19] And again the question will be whether one can keep one's total behavior from involving any net increase to any aspect of any harm.

What I've just said is sketchy, littered with hand-waving "plausible"'s. The details of these cases are complicated and it would take a good deal of empirical and philosophical work to sort them out. We cannot hope to really 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 doing so, however. If what I've argued is right, what we

[18] 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.

[19] An interesting feature of this case is that offsetting is only available under partial compliance with the constraint against unoffset harm increases. One can only decrease the harm if something else will contribute to that harm unless you intervene. We can only reach net zero demand for animal products if nobody purchases meat, in contrast with the CO_2 case, in which we could in principle reach net zero emissions while still emitting some CO_2. So in the animal products case, if everyone were acting as required by the new constraint—purchasing meat only if their purchase is fully offset—then no one would be purchasing meat.

need to look for in these and other cases is not—contra what a theory with a standard constraint against harm would recommend—whether any of our actions will involve doing harm, or whether potential means of offsetting will prevent our actions from harming. 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. But it may be 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 sufficiently offset.[20]

In the remaining sections I raise and address two objections to the account I've offered above. The first worries that I've been too revisionary, the second that, in a way, I haven't been revisionary enough.

7 The Old Provisos Objection

Objection: Suppose we grant you that offsetting cases really involve doing harm, contra the prevention account. This does not yet show that permissible offsetting is incompatible with the kind of constraint against doing harm that many deontologists accept. 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, it is difficult to deny that the punching is harmful yet permissible. Does this refute the advocate of a constraint against harm? No.

[20] Of course, one may have other duties from other sources to do more than this. Besides offsetting one's own emissions, for example, 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 (cf. McPherson (2018)). I think it is important whether the constraint against unoffset harm increases is violated, but it does not fully determine what we may permissibly do.

They have at least two options available.[21] 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 to prevent some much worse outcome.[22] Another option is to say that the constraint against harm does apply in cases like this, but that it, or the reason it provides against doing the action, can be *outweighed* by considerations of general goodness.[23] In either case, if a harmful act is necessary for bringing about some amount of goodness, the act can thereby become permissible. And of course there are all sorts of other provisos to the constraint against harm. Why not think offsetting cases can be subsumed under one of them?

Reply: This strategy is a good one to attempt. It would be ideal if we could capture our intuitions about offsetting cases without revisions to familiar theories. However, as far as I can tell, it will not work. Offsetting has features which makes it unsuitable for assimilating to any of the plausible candidate provisos.

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 for impersonal total goodness. So provisos which say you can 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. This rules out various provisos. And the others won't work either. There is no consent from the victims in the cases we've considered, and the permissibility of increasing harm and fully offsetting it doesn't seem to change even with explicit non-consent from the victims of the harm, so a hypothetical consent proviso

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

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

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

won't work. And as we saw in §3, there need not be any causal connection between the harm decrease and the harm increase it offsets, so we can't assimilate our cases to ones of withdrawing one's own aid.[24]

Perhaps there are other provisos which have been proposed which could explain the permissibility of offsetting. I am not optimistic, however, and think at this point the onus is on the objector to produce an account and show it can get the cases right.

8 The Explanatory Adequacy Objection

Objection: Suppose we grant that replacing a constraint against harm as you suggest does sort the cases in a way that matches our intuitions, allowing harm in offsetting cases but not in cases like Zap Replacement. This isn't a good enough reason to accept your view. We also need some explanation of why the boundary this constraint draws would be a morally significant one. As Frances Kamm puts it, we need to "consider the principle on its own, to see if it expresses some plausible value or conception of the person or relations between persons. This is necessary to justify it as the *correct* principle, one that has normative weight, not one that merely makes all the case judgments cohere" (Kamm 2007, p. 5).[25] You have done no such thing, so your principle remains unjustified.

To make things worse, it's not easy to see what could be the morally relevant difference between doing harm in offsetting cases, which your constraint allows as permissible, and doing harm in what we might call 'replacement cases', like Zap Replacement or Zap Compensation, where the harm one does replaces another harm or lack of benefit.

Zap Compensation

[24] 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.

[25] See Kagan (1989, p. 13) for a similar methodological point.

> *A* zaps *C*, but then gives *C* $200 compensation, leaving *C* at least as well off as they would have been without being zapped or paid.

After all, both offsetting and replacement cases involve doing harm, and they both leave the victim at least as well off as they would have been had one not interfered.

Without some explanation of why there should be a moral difference between these types of cases, we should refrain from adopting a new constraint that would treat them differently. It may be better instead to revise some of our judgements about the cases, either taking both replacement and offsetting cases to be impermissible or taking both to be permissible.[26]

Reply: I agree that revising our judgements about these cases is an option we should take seriously. And I agree that ultimately we should hope that our moral theories will include satisfying explanations of why there are the constraints that there are. And although I will put forward a couple suggestions in a moment, I don't know of an explanation for the constraints I've proposed that is deeply satisfying.

That said, I hesitate to give up on these intuitions and the proposed constraint just yet. We may not have a good account of what moral truth could underlie them, but nor do we have a good error theory that would explain why we'd have these intuitions even though they're mistaken, or a solid argument showing they must be mistaken. We should search for such accounts and arguments, and, once we have some options, compare their merits. It may well turn out that in the end we should treat replacement and offsetting cases as moral equivalents. But it may also turn out that our intuitions about these cases can be vindicated. In the meantime, my own inclination is to tentatively accept the new constraint, since I think it preferable to preserve my judgements about cases until there's fairly strong reason to overturn them. Others may reasonably have other reactions.

[26] Thanks to [redacted] for pressing me on this point.

In any case, like Kamm, I think we should try to give some deeper explanation. But I also agree with her that "[s]ince the principle that is justifiable may be surprising", we should "be prepared to be surprised at what the point of non-consequentialism turns out to be" (Kamm 2007, p. 5). What might the point be behind a constraint that allows offsetting but not replacing harm? I will make two suggestions. One appeals to that familiar deontological heavy lifter: respect for persons and their autonomy. The other suggestion is a more radical one, which makes certain kinds of events—and not just persons—inviolable.

Let's start with the familiar. Suppose we have a case where *A* has a few options: they can pay $200 to prevent *C* from being zapped, they can zap *B* for a $200 reward, or they can do nothing. Many people have the intuition about these kinds of cases that while it would be nice of *A* to pay for *C* not to be zapped, they are not morally required to do so. However, *A* *is* required to refrain from zapping *B*, even though it means forgoing the reward. It seems that there is some morally significant difference between doing and allowing harm. Why should this be?

One attractive answer is that it comes down to the respect for persons' autonomy. 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 ways that they might not want. This is why there is a constraint against doing harm, and why it would be wrong for *A* to zap *B* in this case.[27] Allowing otherwise would fail to respect *B*'s autonomy. On the other hand, a person also needs to be free from certain kinds of normative impositions. Being morally required to intervene on another's behalf is also a kind of imposition, one that morality should respect. This is why there can be no constraint against allowing harms, and why *A* is not

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

required to save *C* from the zap.[28]

I am cutting a long and controversial story short here, since my aim isn't to defend any detailed version of this view, but to see if something along these lines can be used to support a moral distinction between offsetting and replacement cases. If it is to do so, we'll need it to explain two things: why offsetting harm is permissible, and why replacing it isn't.

It seems easy enough to explain why increasing harm and offsetting it would be permissible, according to this picture. Unlike *A*'s zapping *C* for a $200 reward, *A*'s behavior as a whole in Zap Offsetting and Poison 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 difference in welfare for *C*, it's that there's no negative difference at all in the harm *C* suffers, instead only improvements or irrelevant cambridge changes.[29] And since *A*'s not interfering at all would not impose upon *C*, it's very plausible to hold that *C* has not been imposed upon by *A* when *A* increases but fully offsets the harm, either.[30] 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.

The more difficult task is to explain why harm replacement should be impermissible, since it seems we can make a very similar argument for the permissibility of harm replacement. What *A* does in the replacement cases

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

[29] Broome (2012, p. 89) makes a similar point about carbon offsetting.

[30] Note that in making this point we need to be considering the agent's behavior as a whole. Looking at an agent's harm increase in isolation, we might conclude that it is a causal imposition of the sort that needs to be disallowed. So besides appealing to principles of autonomy, we need to appeal to the idea that it's primarily one's behavior as a whole, rather its proper parts, which is to be evaluated. This is close to the idea, recently defended by Portmore (2017) and Brown (2018), that right- and wrong-making features apply non-derivatively only to 'maximal' acts the performance of which are not implied by any other (non-normatively equivalent) acts. But their kind of maximalism isn't quite what is needed here, however, since the acts that compose one's total behavior may may not be unified enough to compose into one big act.

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 very same point could be made about the offsetting case if we single out the harm increasing act.

It seems instead that we should appeal 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". With certain of the replacement cases, this tack seems promising. Even if you should be indifferent between being zapped and being zapped but given $200, it's not too 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 intrinsically identical—not just equal in badness. So how can we say replacing one with the other would be a morally relevant imposition? The thing to say here is that these are different harms—that the pain from being zapped by *A* is not the same harm as that of being zapped by *B*. And even if *C* should be indifferent between them, this doesn't mean that swapping one for the other isn't an imposition upon *C*. The difference in harms allows the autonomy-theorist to say that in Zap Replacement, *A* imposes the harm of the zap on *C*. This harm is something *A* causes and the harm itself is an imposition. In Zap Offsetting, the harm is something that is coming to *C* independently of what *A* does, so is not something which *A* imposes on *C*. So on the autonomy story does seem to provide a rationale for the impermissibility of replacement and the permissibility of offsetting.

We should not be satisfied with this account as it stands, however. For one thing, we should demand a deeper explanation of why we should care about the kind of imposition supposedly involved in Zap Replacement, once we take into account the fact that *A*'s behavior as a whole doesn't seem to change anything for *C* except the identity of the harm *C* undergoes. For another, the basic autonomy-based theory it is based on brings with it plenty of controversial baggage—like the supposed need for protection from normative impositions—which I have not attempted to defend here, and it's not clear to me whether all of it can be defended. However, I do think it is a promising direction to pursue, and there are independent reasons to work out a theory of autonomy and its relations to constraints. It would be a virtue of a theory if it can get our intuitive judgements about offsetting and replacement cases right, and I don't think accomplishing this is a forlorn hope.

So much for my first suggestion. My second suggestion, though it would require a radical departure from familiar deontological ideas, does take inspiration from one of the most familiar: the objection that utilitarianism doesn't adequately recognize the separateness of persons.[31] According to this objection, utilitarians (and perhaps other consequentialists) go wrong in taking it to be irrelevant *who* some amount of harm or benefit goes to, and so allow us to trade off anybody's welfare against anybody else's, so long as we end up with at least as much total good. The objection is that something about the moral status of persons makes such tradeoffs unacceptable, except in special cases.

My suggestion is to locate this kind of moral status, whatever exactly it is, in the harms themselves.[32] On this view, in addition to—or perhaps even instead of—persons, certain events have a moral status which makes trading off an increase to the badness of one with the decrease of the badness of another unacceptable, except in special circumstances.[33] The problem

[31] See Rawls (1971/1999, pp. 26–27), Nagel (1970, p. 134), and Nozick (1974, pp. 32–33). For a recent consequentialist-friendly discussion, see Chappell (2015).

[32] And if we go for one of the aspectual constraints, we'd say it's *aspects* of the harms which have this status.

[33] In diagnosing what we find repugnant about the repugnant conclusion—Nebel (2019, pp. 342–343) suggests we implicitly value the goods in people's lives more than the people themselves, a view similar to the one being considered here. Nebel doesn't himself

with utilitarianism, then, wouldn't (just) be that it is too permissive about tradeoffs of one person's welfare with another's, it's that it is too permissive about tradeoffs of one *harm*'s badness with the badness of other harms or the goodness of other benefits, either inter- or intra-personally. And we would be making the same mistake if we were to allow tradeoffs of one harm with another when they are undergone by the same victim, as in replacement cases. But the kind of 'tradeoff' that offsetting involves, where a harm is increased but also decreased in the same way, is compatible with recognizing this moral status of harms, since no harm (or harm aspect) is worsened. Thus, other things equal, offsetting should be permissible and replacement impermissible.

To be sure, making this move would require a serious rethinking of what this posited moral status amounts to, and what its source could be. For it could no longer be tied directly to some of the most morally salient features of persons, since events do not have those features. It's hard to say what it could be about harms themselves that would confer on them this kind of inviolable status. And until that is done, this proposal doesn't go much beyond the intuition that offsetting is permissible but replacement isn't. It will also exacerbate certain other problems. Why would it be, for example, that tradeoffs among harms often do seem permissible when the victims undergoing these harms consent to it? I cannot offer a serious defense of this separateness of harms view here. Indeed, I am by no means confident that one can be offered. However, I think it is a view worth exploring, and it seems that it could provide an explanation of why there would be the kind of moral difference between offsetting cases and replacement cases that our intuitions about cases suggest.

So now we have two ways our original case judgements and the constraints that capture them might be vindicated with a deeper explanation.

accept this view—indeed, he finds it "morally perverse"—but his argument suggests the alternative is to accept the repugnant conclusion.

But it will have to wait to be seen whether either option—or some other one, yet unconceived—will work out. We should be willing to be surprised.

A final point on this objection: suppose it turns out in the end that we should reject a morally relevant distinction between offsetting and replacement cases. This would mean that the principle I have proposed is incorrect. It doesn't necessarily save the old constraint against harm, though, since it's by no means clear whether the cases would all be impermissible (like replacement cases seem) or else all permissible (like offsetting cases seem). So even if my own proposal ultimately should be rejected, we may still need some other significant revision the the constraint against harm. In any case, more work is called for in understanding offsetting.

9 Conclusion

Sometimes the problem with a moral theory is that it misses a morally relevant distinction entirely, sometimes that it gives great weight to a distinction not actually worth caring about. But often the problem is more subtle: the theory has drawn a distinction which is *close* to a morally important one, but isn't quite carving nature at its 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. I've proposed to instead adopt a theory with a constraint against unoffset harm increases, which gets these cases right, in addition to cases of more ordinary harm. And now I've also outlined a couple options—one radical, the other less so—for explaining why there would be such a constraint, and why the distinction between replacement and offsetting cases would be one that matters.

Only with further investigation will we know whether whether upholding some constraint like the one I've proposed is ultimately tenable. Similarly, more investigation is required to determine how exactly such a constraint would apply to important real-life cases, like carbon offsetting. Given the urgent practical significance of such cases, we should hope that this investigation comes sooner rather than later.

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.

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

Chappell, Richard Yetter (2015). "Value Receptacles". In: *Noûs* 49.2, pp. 322–332.

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.

— (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.

Nagel, Thomas (1970). *The Possibility of Altruism*. Princeton University Press.

Nebel, Jacob M. (2019). “An Intrapersonal Addition Paradox”. In: *Ethics* 129, pp. 309–343.

Nozick, Robert (1974). *Anarchy, State, and Utopia*. Basic Books.

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.

Rawls, John (1971/1999). *A Theory of Justice: Revised Edition*. Original edition published in 1971. The Belknap Press of Harvard University Press.

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 Encycopedia of Philosophy*. Ed. by Edward N. Zalta. Summer 2016. http://plato.stanford.edu/archives/sum2016/entries/doing-allowing/.