Increasing Harm and Offsetting: Replacing the Constraint Against Harm[*]

Mike Deigan

June 13, 2017

The ethics of harming has been much discussed. The ethics of what I'll call *increasing harm*, less so. Behavior increases harm, roughly, if it makes some event more harmful than it would have been, and it *merely* increases harm if it doesn't add any particular harmful aspect or part to the harm that it makes worse.[1] More on what merely increasing harm is presently.

Perhaps the relative neglect of merely increasing harm is justified. It could be that there is no morally interesting difference between merely increasing harm and increasing harm in the more familiar ways of creating new harms or contributing some part of a harm. Views and arguments to do with plain old harm-doing may all carry over to the merely increasing harm case, and a constraint against merely increasing harm might be derivable from some more familiar constraint against doing harm.

If so, moving to the increasing harm case would introduce nothing but needless complexity, at least as far as our normative ethical theories are concerned. In this paper I argue that this is not so. Certain cases of merely increasing harm can be made permissible by a kind of response—offsetting—that is neither available in more familiar harm cases nor re-

[*] Thanks to Dan Ferguson, Yuan Yuan, Zoltán Gendler Szabó, the members of the Fall 2016 Yale Philosophy WIP seminar, and especially to Shelly Kagan for very helpful discussions of this paper.

[1] I will 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 the purposes of this paper 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.

ducible to any of the familiar ways in which harms can be made permissible. We thus cannot derive a constraint against unoffset harm increases from the kind of constraint against harm already widely accepted. In fact, just the reverse: the more familiar constraint against harm can be derived from the more basic constraint against increasing harm without offsetting it.

Increasing harm and offsetting, then, demand our attention both as moral philosophers interested in the content and structure of morality as well as moral agents who often increase harm.[2]

1 Increasing Harm and Offsetting

Considering cases is the easiest way to get a grip on what merely increasing harm is. The interesting day-to-day choices involving what I take to be (risk of) merely increasing harm, like polluting or buying meat, are mindbogglingly complex. For the time being, then, I will limit myself to artificially simple situations in which we are unlikely to ever find ourselves.

Consider, then,

> SINGLE PLAYER, SINGLE VICTIM ZAPPING GAME:
> Perry wakes up in a room which has three buttons: one labelled "Up (+1, +$200)" another labeled "Down (-1, -$200)", and the third labelled "Exit". The room also has a locked door and a screen which displays the name of the game and "Score: 7". Perry realizes he has been abducted and must take part in the Single Player, Single Victim Zapping Game.
>
> In this game, someone is randomly selected to play, and someone else is randomly selected to be the victim. The player is locked in the room until he or she presses the Exit button, which unlocks the door, ending the game and allowing the player to leave. The

[2] The only discussion of offsetting of which I am aware is in an application of the idea to carbon emissions in Broome (2012, Ch. 5). And though Broome doesn't extend the idea of offsetting beyond the case he considers or say how he sees it fitting into any broader ethical framework, and though I have doubts about some important aspects of his view (e.g., the view that preventing others from contributing to a harm, or even preventing others from preventing something (a rainforest, say) from reducing a harm is sufficient for offsetting), it was his insightful remarks in that chapter which brought the idea of offsetting to my attention.

> Up button increases the score and has some effect on the player; in Perry's situation, it increases the score by one and gives Perry a $200 reward. The Down button decreases the score and has some effect on the player; in Perry's situation, it decreases the score by one but costs Perry $200 to press. There is a fixed limit, in this case 10, to the number of button presses the player can make before the Up and Down buttons stop having any effect.
>
> After the game is over, the victim—who in this case turns out to be Vicky—is given a low current, high voltage electric shock for a half-second. How high the voltage is depends on the score of the game: the higher the score, the higher the voltage. And the voltage associated with each score (between 0 and 25, let's say), is perceptibly less painful for the victim than the voltage associated with the score one point higher.

Perry has available various ways to play the game. Here are three: (E) he could simply press Exit, leaving the score at 7; (UE) he could press Up and then Exit, increasing the score to 8 and benefiting himself; or (DE) he could press Down and then Exit, decreasing the score to 6 at a cost to himself.

Assuming that Perry is neither especially wealthy nor especially poor, and that he already does a reasonable amount to promote general goodness, and that the goodness of states of affairs which result from (E), (UE), and (DE) are not greatly different, most of us, I take it, would judge that (UE) is morally forbidden, (E) and (DE) are morally permissible, with (DE) being supererogatory.

As may now be evident, I will be working with a deontological picture of first-order morality that at least roughly approximates common sense morality. I thus assume that while there is a pro tanto duty to promote the general good, there are moderate constraints that sometimes require agents not to 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.[3] For the record, I myself do not have a settled view on what sort of ethical framework, ultimately, we

[3] And while I do take cases of merely increasing harm and offsetting to be less troublesome for garden-variety consequentialists than for deontologists, they do run us into a problem that has been raised for how to coherently formulate consequentialism, as I will discuss in footnote 32.

should be using. I think the best way to find out, though, will be to work out in as much detail as we can the best theories within these frameworks, and compare them to 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. With these methodological remarks out of the way, let's get back to the case at hand.

In explaining why (UE) is forbidden, most of us would appeal to a constraint against doing harm. In pressing U, the thought goes, Perry would be harming someone, and given that his doing so would not be coerced, consented to, an act of self-defense, or necessary to some action which does enough good to outweigh the harm, etc., this is impermissible. So (UE), which has U as a part, is impermissible.

This explanation is plausible. At least if we accept some straightforward counterfactual comparative account of harm, the act U really is harmful to Vicky—she would have been better off had Perry not pressed U.[4] Thus, so

[4] It's less clear if we accept a 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, it seems that this explanation of U's wrongness will not work. For reasons to be discussed shortly, intuitively there is no harmful event which U causes, so it would not count as harming on this theory. A more sophisticated version of the theory, such as the one Harman (2009) ends up with, can count U as a case of harming. The trick is to treat harms as being extremely modally fragile, so that causing any difference causes a different harm. This means pressing U does cause the harm Vicky suffers. Then to prevent overgeneration (to keep, say, a pressing of D from counting as harming Vicky), we restrict 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. U 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 pressing of U as a case of harming (it is part of the sequence leading to the harm to Vicky), but as it stands also would treat the pressing of D as a case of harming, which is unacceptable. But the account could be easily modified along the lines which Harman modifies the simple causing harm account, by treating only worsenings of harms as harmings. This would then predict that U is a case of harming.

From here on I will ignore the differences between these and other accounts of doing harm, since none of the differences will matter to the arguments in the paper, as far as I long as it doesn't meet some criterion for permissible infringement of that constraint (and it doesn't seem to), this constraint explains why it's wrong.

But this account elides an important aspect of the case: unlike familiar cases of doing harm, the harming involved in pressing U is one of merely increasing harm. Usually when we harm someone, there is some event which is intrinsically harmful[5] to them that occurs that wouldn't have otherwise occurred, or at least some particular intrinsically harmful part of that event that wouldn't have been there.[6] If someone zaps you with a taser, there's some intrinsically harmful event—the occurrence of the pain of being zapped, say—that would not have occurred had they not zapped you. And if two people each shoot one bullet into a crowd, there are intrinsically harmful events that wouldn't have happened that come from each shot, even if we don't and cannot know which shooter produced which harm. But in the case of (UE), there is no such connection, even in principle. The pain from the zap does not have parts which can be divided up into those ascribable to Perry and those not.[7] And it would have occurred—though it wouldn't have been as bad—even if Perry had

can see. All the accounts that can explain (UE)'s wrongness in terms of U's being a case of doing harm will fail to accommodate offsetting.

[5] An event is intrinsically harmful to someone if that event, in itself or constitutively, contributes negatively to that person's welfare. This is meant to be analogous to the more familiar notion of intrinsic value or disvalue. And it's meant to be analogous to intrinsic value 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. Bradley (2012) makes use of 'intrinsic harm', but conflates these two senses.

[6] We might wish to reformulate this claim in order to capture overdetermined harms, though whether we would have to do so may again depend on our views of how to evaluate counterfactuals, about the essences of events, as well as what we want to say about the morality of overdetermined harms. Similarly for harms—if we want to treat them as harms—of pre-emption or omission.

[7] There are various ways we might conceive of this pain and how it can be divided into parts. If it's identical to a mental event, for example, we might take it to have spatio-temporal parts such as this or that neuron's firing. Or we might go in for a somewhat extended notion of part (c.f. Paul (2002)) to allow that things such as the pain's duration or its intensity will count as parts. Less controversially, I assume we can take it to have temporal parts, such as the first quarter second of the pain. distinguish aspects of the pain (such as its temporal length and its intensity). But on any reasonable view of the pain's parts, none of the morally relevant ones are fully attributable to Perry.

instead done (E).[8] This is what makes (UE) and the U part of it, a mere increase in harm.

This may be so, of course, without there having been anything wrong in our first, flatfooted account of (UE)'s wrongness. Any case of harm will have its idiosyncrasies; this doesn't mean an account that doesn't mention them is missing something morally relevant. What, then, is supposed to be missing for our case?

What is missing can be brought out by considering our reactions to some of Perry's other options beyond the three we've mentioned so far. Here are a few: (DUE) Perry presses Down, then Up, then Exit; (UUUDDDE) Up three times, then Down three times, then Exit; and (DUUDE) Down, then Up twice, then Down and Exit. What all these have in common is that each of their net effects on the score are neutral. In addition to mere harm increases, they also have mere harm decreases that offset the increases. And, apparently because of this, they seem to all be permissible.[9]

But in each of these permissible options, there is at least one U. So if the occurrence of the U in (UE) is what made it wrong, why aren't (DUE), (UUUDDDE), and (DUUDE) impermissible due to the U's that occur in them?[10]

The answer is that the U's in the permissible cases are sufficiently offset by D's. What makes (UE) impermissible is not that there is some part of it that is a harm which fails to be justified in the usual ways, but rather that

[8] To say that the event of Vicky's feeling pain would have occurred even if Perry hadn't made it worse does 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. This view seems false to me (for an argument against very fragile events, see Lewis (1986, pp. 196–199)), but I don't wish to discuss it here, since I take it that we could carry out the same discussion in different terms without the assumption that it is false. We would just need to introduce a counterpart relation that relates the events that I would take to be the same, and say that the intrinsically harmful event caused by (UE) has an intrinsically harmful counterpart in the worlds where Perry does (E). We need not quibble, at least at this point, over whether these events are *really* the same.

[9] Permissible even if pointless time-wasting. This is no problem, as most of us accept that so long as it's our own time that's being wasted, it's often okay to do pointless, time-wasting things. But in case making it not pointless makes its permissibility a little clearer, suppose Perry also gets $1 each time he presses a button, then some combination of 5 Ups and 5 Downs would be a sensible, permissible think for Perry to do.

[10] A related question around which this paper could have been framed: in the case of (E), Perry did not press D, but did not fail to do anything required of him, so why is it that it is required for him to press D if he has pressed U?

the mere increase to harm it involved was not and would not be fully offset by Perry. This is what the original account missed. And this is why we need to introduce a new constraint against unoffset harm increases, rather than trying to explain (UE)'s wrongness by appeal to a constraint against doing harm which we already have.

Or so I say. We might worry that I have moved too quickly. Implicit in the original account was the claim that the familiar reasons for lifting or overriding the constraint against harm are not present in (UE), so if we can show that some of them are present in the offsetting cases, then our original account might not have been off the mark after all. This worry is a serious one, so I will address it at length in the next section, arguing that attempts to assimilate offsetting to other, familiar ways of negating or overriding the wrongness of a harm fail.

2 Against Offsetting Skepticism

What should those who doubt there's anything wrong with our original explanation of the impermissibility—we'll call them the offsetting skeptics—say to make sense of the permissibility of (DUE), (UUUDDDE), (DUUDE), and other U-involving net neutral possibilities available to Perry?

They could, of course, just deny that these latter options really are permissible. Without some further argument, though, this is move seems neither very interesting nor very plausible. In considering the case, these options do *seem*, at least to me, to be permissible. The skeptic would have to assert that this intuition isn't tracking the truth. And though I can't refute such a claim, I do think the following kind of thought places the burden on the offsetting skeptic to justify it. We all agree, I take it, that (E) is permissible. Yet there is no difference for Vicky, the only person besides Perry who might be harmed or benefited by his actions in the zapping game, between (E) on the one hand and (DUE), (UUUDDDE), and other net neutral cases on the other.[11] Moreover, the general good resulting from (E) is no higher than that which results from the net neutral

[11] Note that it's not merely that she's not made worse off on the whole. As we will discuss, complex actions combining harms and benefits might still be wrong even if they don't leave the victim worse off on the whole. Rather, there's no difference at all, irrelevant Cambridge changes aside. The harm that comes to her in any of these cases will be a zap with the voltage associated with the score of 7.

offsetting cases. So how could they but not (E) be wrong? It might turn out that there are wrong-making features that are neither person-affecting nor general-goodness-affecting, but without some story about what they are and why they rule out (DUE) and other offsetting cases but not (E), appealing to them is ad hoc.

A more interesting and plausible strategy for the offsetting skeptics is to accept the judgments about the cases but appeal to some familiar explanation of why the offsetting cases are permissible, perhaps by showing that the harms that they involve are permissible ones. Some of the moderately well-understood ways in which harms can be made permissible, like the harm's being done as part of proportionate self-defense, are obvious non-starters. Others have more to be said for them. It is these more promising candidates that I will be arguing against for the remainder of the section.

2.1 Moral Balancing

Practically nobody thinks that every possible harm to an innocent, non-threatening victim is impermissible. We can imagine cases which force a trade-off between harming an innocent person and allowing many more people to be harmed in a much worse way. If, somehow, the only way to prevent 1,000 people from being killed is to punch some innocent person, it would be crazy not to think that punching the innocent is harmful yet permissible.

Does this refute the advocate of a constraint against harm? No. She has at least two options available.[12] 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 preventing some much better outcome.[13] 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 other considerations.[14] In either case, once some threshold of goodness a harmful act would be involved in bringing about (or other moral reason-giving considerations) is passed, it

[12] See, e.g., the discussion of the scope of and thresholds for a constraint against harm in Kagan (1998a, §§ 3.2–3.3).

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

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

can become permissible. I will call these two approaches to explaining a harm's permissibility *moral balancing* approaches.

Can the offsetting skeptic explain the permissibility of offsetting cases by appealing to moral balancing? I think not. What they would have to say is that while the U in (UE) was unnecessary for bringing about something sufficiently important to balance out the harm it does, the U's in the offsetting cases *are* required to bring about something that is sufficiently important to balance out the harm they do. But what might that be?

The U's in the offsetting cases seem to be doing exactly what the U in (UE) is doing: increasing harm to Vicky and benefiting Perry. It's not as if the U's in the offsetting cases increase the score by any less, or benefit Perry any more, than the one in (UE). Each one increases the score by 1 and gives Perry $200. And it's not as if they are required for some larger action which brings about a great amount of good. We can imagine cases where an increase in harm opens up possibilities to do something very good. For example, we can imagine a case where pressing U will make available another button for Perry to press which would cut the score in half, or prevent the game from ever being played again. But nothing like this is going on with the offsetting cases we've been considering. If anything, pressing a U restricts Perry's options to do good, including in the offsetting cases. Given the limitation on button presses, he can now press D one less time than he could before.

The offsetting skeptic might try saying that while in (UE) the U is merely giving some gain to Perry, in the offsetting cases it is preventing losses, losses that will come to Perry from having pressed D's. And preventing losses, the offsetting skeptic could claim, is important enough to balance out the harm of pressing U.[15]

Depending on how the payment to press a D is made, this suggestion might help with certain cases. Suppose, for example, that the loss is delayed. A balance is kept, and when Perry presses E, the balance is deposited or withdrawn from Perry's bank account. Now take the U in (DUE); the situation Perry faces just before he presses U is a loss of $200, leaving the score as it currently is (now 6), or preventing that loss by increasing harm to Vicky. It's not obvious that preventing this kind of loss really would be enough to balance out harming Vicky, but it's also not obvious that it couldn't. And at any rate, this situation is different from

[15] Perhaps the self-defense suggestion isn't so obviously false after all.

the one that Perry faces when he decides to press U in (UE), where there is no impending loss to Perry and the score is 7.

This strategy, however, depends on too many idiosyncrasies of this case to be promising as a general strategy for dealing with permissible offsetting. For one thing, we don't have to say that the loss to Perry for pressing D are delayed; perhaps he has to pay in cash each time he presses it, or, if we like, we can modify the case to make the cost to press D not monetary at all, but instead, say, a small zap to the player at the moment he presses it. Then when he's deciding whether to press U after having pressed D, there's no impending loss for Perry to be preventing.

Moreover, regardless of when the loss comes, it's unclear how the loss-prevention moral balancing account could make sense of why offsetting cases which start with a U, like (UUUDDDE), are permissible, since there is no impending loss for Perry to be preventing when he presses U's in these cases. Perhaps we could say that the U's are only permissible in these cases when Perry plans on pressing D's later to offset them, and so in a sense there are impending losses for him to prevent. But for those who favor a constraint against harm, I take it, preventing this kind of foreseen, freely and deliberately self-imposed losses is not going to be the sort of thing that can morally balance a harm and make it permissible.[16]

Finally, I think there are cases of permissible offsetting for which there is no hope at all that this strategy will work. Imagine a variant on the zapping game where neither U nor D costs or benefits Perry, but D can only be pressed when preceded by a U. In such a situation, it seems, (UDE) and other options involving U's fully offset by D's are permissible. But here there is no relevant loss for U to prevent, and no way for U to prevent any loss there might be, so the loss-prevention balancing account can't be explaining why the U's are permissible. And more generally, it's hard to see how a moral balancing account could make sense of this sort of

[16] Indeed, I think considerations along these lines make it difficult to see how the loss-prevention account could make sense of even the cases we started with, for which it seemed most plausible. It's not just any loss they would be preventing, but a freely and deliberately self-imposed loss. Perhaps I can harm an innocent who, if they are not harmed, will impose some loss on me—perhaps, though this seems less likely, I can even harm an innocent bystander to prevent some impending loss to me—but it is pretty clear that I can't harm them to prevent some loss to me if I am responsible for the threat that they pose. I think a better approach which is more likely to capture what, if anything, is right about this moral balancing account, is to view what's going on as withdrawing aid. But this too, as we'll see in §2.4, also runs into some of the same problems.

permissible offsetting case. The increases of harms are only necessary to decrease harm by that same amount. Nothing good comes from this that wouldn't come from (E).

The moral balancing proposal, I conclude, doesn't explain why offset harm increases are permissible.

2.2 Compensation

The offsetting skeptic might hope to make progress by making the following observation: it seems important that the benefits of the offsetting D's are benefits that go to the victim of the U's. If D reduced the score of some other, simultaneously played zapping game, whose victim was Victor rather than Vicky, Perry could not permissibly play (DUE), even if the D reduced the score for Victor by 2 or 3 points. Leveraging this observation, we could say that it's not some general, interpersonal moral balancing that's going on in the offsetting cases, but rather sufficiently compensated harm.[17] Is a compensation account able to make sense of offsetting? I will argue that it isn't.

We can divide compensation accounts into broadly two camps, which I'll call trade-off accounts and making-up-for accounts. The former limit the constraint against harm to those that don't lead to a sufficient amount of good for the victim of the harm.[18] Harms sufficiently compensated for are not, on this kind of account, wrong. The making-up-for accounts are stranger: they say that some wrongful harms can be fully made up for through compensation, and thereby be involved in complex actions which are not, as wholes, impermissible. Let's take these two kind of accounts in turn.

The straightforward trade-off accounts run into the problems similar to those raised for as the general moral balancing accounts. First, there's the

[17] See Parfit (1984, §115) for a helpful discussion of the distinction between moral balancing and compensation. One prominent use of compensation to temper constraints is Nozick (1974, Ch. 4).

[18] Alternatively, as with moral balancing, they can keep a simple constraint but say it can be outweighed by doing enough good to the victim. Kagan (1998a, p. 88), for example, makes this suggestion. Perhaps there could be some threshold, which, if surpassed by some weighted combination of general good and good done for the victim (with the latter receiving more weight), allows the constraint against harm to be outweighed. Whether the trade-off idea is cashed out in terms of threshold or complications to the content of the constraint is not, as far as I can see, relevant to the issues I will be discussing.

problem of saying what compensating good for Vicky comes from Perry's pressing U in the offsetting cases. Second, if we are somehow able to solve the first problem by appealing to the good for Vicky that comes from the offsetting D's, we will still have trouble with making sense of cases, like (DUE), where the offsetting comes before the harm increase. Finally, there seem to be permissible offsetting cases which any strategy along these lines will fail to account for.

Take the first problem. Perry plays (UDE). On a trade-off compensation account, this is permissible because, although U is harmful, it is fully compensated for. What is it compensated by? Presumably by the equal but opposite score change effected by D. I find this theory hard to accept for two reasons.

First, normally when we take a case of harm to be permissible because of the compensating goods that come to the victim, the ratio of the harm to the compensating good is not 1:1. Harms done to bring about equally-sized goods for the victim generally aren't permissible. I may be able to amputate your leg without your permission (or perhaps even against your will in some cases) in order to save your life. But, assuming that they are of equal value to you, I cannot (without your permission) amputate your right leg to save your left leg (which would be lost painlessly). The harm-to-benefit ratio in the offsetting cases, though, are 1:1.

Second, in the normal permissible compensation cases, the harms have to be *required* (or at least, some harm that is at least as bad has to be required) to bring about the compensating good. I can't permissibly amputate your leg for no reason but my own gain if I am going to save your life from some unrelated threat afterwards.[19]

Even if we are able to find responses to these worries, it is hard to see how they would work for cases where the 'compensation' comes first, as in (DUE). I can't save your life and permissibly follow it up by amputating your leg if it doesn't further benefit you.

Suppose, however, that we are able to find some account that satisfactorily avoids these issues for (UDE) and (DUE) in the normal zapping game. There are other cases of permissible offsetting which, it seems to me, for which this kind of account is useless.

[19] Perhaps I can permissibly demand the trade: let me take your leg and I'll save your life. But if you don't accept the trade, clearly it would be impermissible for me to amputate your leg anyways, even if it would be rational for you to have accepted to deal.

Consider the Single Player, Multiple Victim Zapping Game. In this game, the score is not correlated with the voltage of the zap, but rather the number of people who are to be zapped. At the end of the game, a number of victims equal to the score are randomly selected and are all given a moderately painful zap. Offsetting in such a game seems permissible, but cannot be explained in terms of compensation to the victim(s). Leaving the score as it is does nothing to compensate those who are in fact zapped, as they are zapped just as badly.

Now let's briefly consider the making-up-for version of the compensation idea. On this view, (UDE) is permissible even though the pressing of U that it involves is wrong. The proponent of this view could point to cases where, having wronged someone, we are obligated to compensate them, and that once we've done so it seems that in some sense everything is okay again. They would account for this by saying that while the original act was wrong, the whole, complex act of wronging and compensating is permissible, whereas complex acts of wronging and not compensating are impermissible.[20]

Inherent in this view is a denial of the very plausible principle that impermissibility of an action is inherited by the larger actions of which it is a part.[21] So it has to start off by biting a bullet, which I think counts against it as an account of anything, offsetting included. But we all must bite bullets at some point, and I think the view that (actual) wrongness of parts need not imply wrongness of wholes is an interesting view that must deserve more attention than it usually gets. So I suggest we look past this initial implausibility of the making-up-for version of the compensation account and see if it helps with offsetting.

And in one way, at least, it does better than the trade-off account. For unlike that account, it doesn't run into the second problem about the harm of U not being required for the benefit that comes from D. In order for the whole action to be permissible, the wrong part has to be made up for by the compensation, but the wrong part doesn't itself have to produce (or be necessary for producing) the compensation.

But the making-up-for account still faces the other problems. I haven't

[20] Note if this is right, we would have to make a revision to our original account of the wrongness of (UE). We can't just say that U was wrong so (UE) was wrong; we have to add that it was an uncompensated wrong, so (UE) was wrong. But this account, I think, would support the offsetting skeptic's project in spirit, if not in letter.

[21] Cf. the Means Principle for Permissibility from Thomson (1990, p. 112).

sufficiently made up for some harm done to you if I merely do something that makes you no worse off than if the harm hadn't happened. And this account has, I think, an even worse problem with the reversed order cases, where the 'compensation' would be coming prior to the harm. We cannot normally make up for some wrong that we haven't yet done. Putting these problems together in a single case: if I prevent your right leg from being wrongly amputated, I can't then amputate your left leg for no reason, then say that the whole, complex action was permissible because while I did wrong you by amputating your leg, I sufficiently compensated you for it beforehand by saving your other leg. And this version of the compensation account does no better than the trade-off account in making sense of offsetting in the Multiple Victim zapping game. So I don't think the making-up-for compensation theory is a good way of making sense of offsetting.

The compensation proposal doesn't explain why offset harm increases are permissible.

2.3 Hypothetical Consent

The offsetting skeptic need not (yet) despair, for there are familiar cases of harm which seem permissible yet don't seem to be compensated or balanced out by morally important gains. For example, suppose you want to test out your new taser. You ask me if I'd like to volunteer to be tazed. Uncoerced, fully informed, and of sound enough mind, I consent to your tazing me. And you do it. Being tazed hurts a lot, and so harms me. And there doesn't seem to be anything important enough here to outweigh the badness of violating a constraint against harming. But nevertheless many of us find the tazing in this case permissible, so it seems that something about consent can keep harm from being impermissible. Offsetting skeptics may look to consent as a way to explain why the offsetting cases are permissible.

As there's no mention of consent from the victim in our original description of the zapping game, though, we can't say that it's the presence of actual consent in the offsetting and its absence in the non-offset harm increase cases that makes the difference. They may, though, appeal to *hypothetical* consent.[22] In the offsetting cases, they could say, Vicky *would*

[22] See Kagan (1998a, pp. 88–90) for a discussion of consent and the move to hypothetical

consent if she were asked (and was being reasonable and not being coerced, etc.), and that this is enough to make the harm in those cases permissible. She would not consent, however, to (UE), so the harm it involves is wrong, as we said at the outset.

Like many, I have my doubts about the non-derivative moral significance of hypothetical consent. But putting these doubts aside, I still don't think appealing to hypothetical consent is a good way for the offsetting skeptic to explain the permissibility of offset harms. I will raise two problems for doing so.

The first is that so far nothing has been said about *why* Vicky would consent to the offset harm increases but not the unoffset ones, and the obvious answer—"Because they are offset"—gives up the offsetting skeptic's game. Unless the skeptic can give some account of the reasonableness of consenting to offset harms which doesn't appeal to offsetting without explaining why it matters in other terms, we have no reason to think that the offsetting itself isn't doing work in explaining why offset harms aren't wrong. This isn't just the general worry about how hypothetical consent could have any non-derivative moral force; rather, it's that even if it is doing some work, we don't yet have a story about why there would be hypothetical consent that doesn't rely on an appeal to offsetting's normative significance. We haven't yet ruled out that no non-offsetting-based story can be told—though we have already seen reasons why some ways of going about it, like those which would appeal to some kind of balancing or compensation, won't work—but the appeal to consent will have to go beyond a simple assertion that there would be consent in order to be useful to the offsetting skeptic.

A second, more concrete problem with appealing to hypothetical consent is this: on any plausible view, in the face explicit non-consent (that is reasonable, informed, uncoerced, etc.), hypothetical consent is either impossible or morally null.[23] But we can imagine cases of offsetting in the face of explicit non-consent that seem permissible. Before the game

consent.

[23] Whether it's impossible will depend on what hypothetical consent is supposed to be and what we think about 'counterfactuals' with true antecedents. If hypothetical consent simply is the truth of the counterfactual "If she were asked (and were reasonable, etc.), she would consent.", and we think that a counterfactual must be false if the antecedent actually holds and the consequent doesn't, then hypothetical consent is impossible if she is asked (and is reasonable, etc.) but does not consent.

started, Vicky might have publicly announced that if she is ever chosen as the victim, she denies consent to press U even if it has or will be offset by a pressing of a D. But this seems to have no more effect on the permissibility of (DUE) than would her announcing that she consents to no play which doesn't decrease the score by at least one point has on (E). Which is to say: at most very little. So long as he sufficiently offsets all of his U's, leaving the score no worse than when he started playing, how Perry decides to play is just not the sort of thing for which Vicky's consent or non-consent is especially relevant. "It's none of her business", Perry might reasonably say.

The consent proposal doesn't explain why offset harm increases are permissible.

2.4 Withdrawing Aid

Cases of an agent withdrawing aid which the agent herself provided bear some similarity to offsetting and have received a fair amount of attention.[24] Maybe the offsetting skeptic can assimilate offsetting to them.

Suppose you see that someone is falling from the window of a burning building, but there's a net positioned underneath them that will keep them from suffering more than minor injuries. But now you see that two people are falling from the next window over and will land a few feet to the right of where the net currently is, and so will be badly hurt by the fall. Is it permissible for you to move the net under the two people, so that they will each sustain only minor injuries, allowing the one to suffer very bad injuries?

Faced with this description, many of us would say "No, it's not permissible," since apparently it would violate a constraint against doing harm.[25] But suppose we add more to the description of the case, giving us two variants. In one variant, the net had been placed there by a fireman in order to save the first falling victim, and now the fireman is off doing something

[24] For discussion of withdrawing aid, including cases similar to those I will be considering, see McMahan (1993) and references therein, as well as Woollard and Howard-Snyder (2016, §7) for an overview of the more recent literature. McMahan is primarily concerned with the killing/letting die distinction, but most of the considerations can be extended to more general distinctions, like that between doing and allowing harm.

[25] Though philosophers disagree about whether this really is a case of doing harm. See Woollard and Howard-Snyder (2016, §7).

else. Here, I take it, most people would maintain that it is wrong to move the net. In the second variant, you yourself placed the net there to save the victim. Here it's less clear that it is wrong to move the net. Suppose we have a principle, then, that implies that it takes a much stronger reason to make permissible a normal harm or a harmful withdrawal of aid that someone else has provided than it does to make permissible a harmful withdrawal of aid that oneself has provided.

Then the offsetting skeptic could try to appeal to this principle in order to make sense of the permissibility of U's in the offsetting cases. The idea would be that, e.g., in (UE) the U is a harm, and cannot be made permissible by the $200 gain to Perry. In (DUE), however, the U is not simply a harm, it is a harmful withdrawal of the aid Perry provided with D. Thus, even though $200 would not be able to make the U in (UE) permissible, it might still be that it is enough to make the U in (DUE) permissible.

Despite its initial plausibility, this analysis faces a number of problems. First, it's not at all clear how it could be extended to cases where the offsetting comes first, like (UDE)—how can you withdraw aid you've provided before you've provided it?—so it seems at best a partial solution.[26] But I don't think it will succeed even in accounting for cases with the right order.

One problem is it's difficult to make sense of the idea that U in (DUE) is withdrawing something that you have provided. Sure, it adds the same amount to the score that D subtracted, but does it make sense to say that there was some separable thing (a -1 point?) that Perry put out there as an aid to Vicky and now he's taking it back? In (DUU), is just one of the U's wrong? And if so, which one?

Even in cases where there are identifiable things we can add and take away in offsetting, the offsetter is not required, as the withdrawing aid account would have it, to take away what he or she has contributed rather than what was already there. For example, suppose that rather than buttons, there's a pot in the room that has tokens in it, and the score is determined by subtracting the number of D-tokens in the pot from 20. When Perry starts playing, there are 13 D-tokens in the pot, so the score is 7. Perry is given 5 D-tokens, and it costs him $200 each to add them to the pot. And he is given $200 each time he takes a D-token from the pot. Suppose Perry

[26] And even if we could make sense of this as withdrawing aid, we couldn't justify in cases like the one we saw at the end of section 2.1, for which U doesn't benefit Perry but makes D available.

is going to add a D-token to the pot and then take one out, leaving the score at 7. Is the complex action permissible only if he makes sure to retrieve the very same D-token from the pot which he adds? It appears to make no difference. Thus it seems that the withdrawing aid account fails to make sense of offsetting, even in the cases most favorable to it.

The withdrawing aid proposal doesn't explain why offset harm increases are permissible.

2.5 Prevention of Your Would-Be Harm

Another account that might appeal to the offsetting skeptic involves an idea similar to withdrawing aid, but denies one assumption we've been taking for granted until now. It denies that when U's are offset they are really harms. What offsetting does, on this view, is prevent an action that would have been harmful from being harmful, thus making the joint performance of the actions permissible.[27] For example, what the D in (UDE) does is prevent the U from doing any harm, so that U doesn't violate the constraint against doing harm and so, given that nothing else makes it wrong, isn't wrong.

There are indeed many cases where we can permissibly do something that would be harmful were it not for some other action we in fact go on to perform. Cases where we lead someone to rely on us often have this structure. If for one of those trust-building exercises I promise to someone that I'll catch him when he falls backwards, and he lets himself fall because of this promise, my promising harmed him if I don't in fact catch him.[28] Catching him prevents me from harming him, and so makes permissible what would have otherwise been impermissible.

[27] For recent discussion of how deontologists should treat actions which can change the status of past actions in this way, see Hanna (2015a) and Hanna (2015b). As he makes clear, there are difficulties in doing so. Nevertheless, the offsetting skeptic may hope that they don't require fundamental rethinking of a constraint against harming.

[28] Perhaps something like this thought could figure into an account of the wrongness of promise-breaking. It's not so much that the breaking of the promise harms the person (it usually doesn't any more than other people's failing to do what the promiser promised to do), but rather that the promising leads someone to do something that will harm themselves unless the promise is kept. Breaking a promise, then, is sometimes a failure to prevent your own harm. I don't mean to suggest this as a complete account of the wrongness of promise-breaking.

We might think that a prevention account can't make sense of offsetting that comes before the would-be harmful act; how can you prevent the harm of something that hasn't happened yet? But in fact there are plenty of prevention cases that have this order. 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 unloaded the gun after picking it up, so it wasn't harmful. So the prevention account might work for explaining the permissibility of (DUE) in this way. The D makes U into an act that doesn't harm.

It can feel odd to say that some action which adds to the score, given that a higher score makes the zap more harmful, doesn't in fact do any harm, but there does seem to be something right about it. After all, one of the thoughts we've been appealing to is that when the U's are offset, there's nothing at all different for Vicky than if Perry had pressed (E), so how could she have been harmed by them?

The prevention of would-be harm account, then, seems a promising one for the offsetting skeptic to adopt. But, like the other accounts we've examined, it fails.

The main problem it faces is that it is important to the harm prevention account, as it was to the withdrawing aid account, that the offsetting be linked to the particular would-be harmful acts that the offsetter performed. But in most permissible offsetting cases it's irrelevant or impossible for the offsetting acts to be linked in this way to the would-be harmful ones.

Take the normal zapping game cases. Are scores the kind of things that can have particular parts taken out of them by the score being reduced? And if so (which is implausible), how do we know that the offsetting has taken out the right parts? Having pressed U, increasing the score to 8, how can Perry be sure that it's the very point that he added that's being subtracted by pressing D? And when he starts off by pressing D, how is this not subtracting some other point, rather than preventing his pressing U from being harmful?

And cases where there do seem to be identifiable contributions, it doesn't seem to matter that they are the ones which are offset. Consider another token version of the zapping game. This time, the score is determined by the number of U-tokens in the pot. Perry starts out with 5 U-tokens of his own, and he is paid $200 each time he adds a token to the pot, and must pay $200 to take one out. At the start of the game, there are 7 U-tokens in the pot, and Perry's first action is to add one of his U-tokens.

But then he pays $200 to offset his addition by taking a U-token out. Must he ensure that the one he retrieves is the same one he takes out? Surely not. But if he retrieves a different token, then his offsetting doesn't prevent his first action from increasing harm. The token he added really in the end increases the score and so increases the harm to Vicky.

The prevention of would-be harm proposal doesn't explain why offset harm increases are permissible.

2.6 Blocking Harm Enablers

There's a final proposal I'd like to consider on behalf of the offsetting skeptic.[29] It's much like the prevention of would-be harm proposal, in that it denies that the U's in offsetting cases harm (or increase harm). But unlike that proposal, it also denies that U's do or increase harm in the non-offsetting cases like (UE). On this view, U's themselves are always harmless; it's some other action, such as E or the whole complex action (UE), which increases harm in the unoffset cases.[30] And the corresponding action in the offsetting cases, such as (DUE) or the E it contains, does not increase harm. So on this view, it's not that offsetting prevents U from being harmful, but rather that the U in unoffset cases makes something else harmful, whereas something about offsetting prevents this from being so. To put it briefly: offsetting blocks an enabler of harm.

I think this proposal is an interesting one with intuitive appeal, but that ultimately it won't work. First I'll argue that in the zapping game cases we've been considering, E's don't increase harm, and that the complete action (UE) doesn't increase harm unless the UE part of the offsetting case (DUE) does too. Then I'll introduce a variant on the zapping game for which the proposal more clearly fails.

I assume it's agreed that the E in (E) does not increase harm, so if we want to say E harms in the (UE) case, we'd better also say the E's in these cases do something different. It's not a brute fact that one harms and the other doesn't. In the original case we can in some sense differentiate what each E does to the zapping setup, since the setup is in a different state: for (UE) the score is at 8 and 9 button pushes remain when E is pressed, for

[29] It was first suggested to me by Zoltán Szabó, though several others have had reactions along these lines.

[30] So this too deviates slightly the original proposal for explaining the wrongness of (UE), but in a way that preserves the main idea that there's nothing new offsetting.

(E) the score is at 7 and 10 button pushes remain. So at least in terms of what states they bring the setup from and to, these actions differ. So far, so good.

Such differences, though, cannot be the source of a difference in harmfulness. To see this, consider a game where the score starts at 8 rather than 7 (and where the maximum number of non-E button pushes is 9 rather than 10). Here too we want to say the E of (E) doesn't harm. But it could be that, at beginning of this new game, the state of the machine of the game and everything in the causal pathway from it to the intrinsically harmful event (Vicky's pain) is an exact metaphysically intrinsic duplicate of the state of the machine, etc., from the original game after the player presses U. But whether some button-pushing increases harm can only depend on these intrinsic features, so if E does harm in (UE) for the first game, then it does harm in (E) of the second. But it doesn't in the latter, so it doesn't in the former. So it can't be E in (UE) that increases harm. The very same reasoning can be carried over to show why (UE) as a whole can't be increasing harm in (UE) but not in (DUE): (UE) in the original setup may be an exact duplicate of the UE in (DUE) in the modified setup, so if it doesn't increase harm in the latter, it doesn't in the former.

Of course, there are extrinsic, past differences between these events, but they are only differences in what led to the present state of the game. And while these may be morally relevant, I don't think they can be relevant to whether some button press harms or not, since they don't affect any of the intrinsic features of the pathway that leads from the button press to the harmful event. Compare this to the gun example. Whether pulling the trigger harms depends on whether the gun is loaded, not at all on who loaded it. Whether the pressing of E harms, similarly, shouldn't at all depend on who presses it or what they've done in the past, except insofar as what they've done alters something about the pathway from E to the pain of the zap.[31]

The blocking harm enablers strategy, I conclude, does not explain what appear to be permissible offset harm increases.

But in case the reader is unconvinced that this strategy fails to account

[31] Perhaps we can raise doubts about this causal-effects-supervene-on-intrinsic-feature-of-causal-path principle, but it seems like a good principle to me—and not just to me: see Paul and Hall (2013), especially Ch. 3 §4.3, for an elaboration and qualified defense of the intuitive idea of causal relations as intrinsic. I'd want first to see a good case against the principle, as well as reason to think it doesn't apply in this case.

for our original zapping game case, I now introduce a variant for which it more clearly fails to account for our intuitions about permissible offsetting.

> U-Rocks and D-Balloons:
> This time there are no buttons. Perry may leave when he wishes, but only has 5 minutes before the zapping happens. How high the voltage is will depend on the weight on a scale in front of him at the moment just before the zapping; the greater the weight, the greater the voltage (a barrier is removed and the scale depresses a pump which generates the electricity). Perry can adjust the weight on the scale by adding 1 lb U-rocks, for which he is rewarded $200 each, or by tying on -1 lb D-balloons, for which he must pay $200 each. Currently there are 7 U-rocks on the scale and no balloons.

Now we want to consider (U), where Perry adds a U-rock and doesn't otherwise affect the scale within the 5 minutes, (UD), where he adds a U-rock and a D-balloon, and (), where he adds neither. With (U) Perry increases harm. I say this is because the U increases harm, just as it does in (DU). But on the blocking harm enablers proposal, we must deny this.

This is bad. It seems clear that the U causes a harm increase in (U) or any other action Perry might have performed that included it without preventing the harm altogether (by adding at least 8 D-balloons). There's a rather direct causal path we can follow from the U-rock placement to the greater harm. Perry puts the rock there, it exerts downward force on the scale, which at the relevant moment contributes to higher voltage just like the other rocks, which makes for worse pain (and note that it does so whether or not he then adds a D-balloon; the balloon doesn't affect the force of the rock Perry placed, it just exerts its own oppositely directed force of equal amount). How could this not increase harm? I don't see a way of plausibly denying that it does. And what else in (U) would increase harm? I see no plausible alternative. And once we admit that the U does increase harm, we see that the blocking harm enablers strategy cannot work to account for the permissibility of offsetting in general, since just as in the zapping game, actions with offset harm increases like (DU), (UUUDDD), (DUUD) and seem to be just as permissible as ().

2.7 Conclusion

We've now considered and rejected six ways for the offsetting skeptic to explain the permissibility of offsetting mere harm increases in already familiar terms. Perhaps there are alternatives which I have overlooked, or perhaps some complicated amalgam of the accounts we've considered could be made to work. But I am out of suggestions for the offsetting skeptics, except for this one: tentatively give up offsetting skepticism, and explore what it would look like to add into one's theory a constraint against non-offset harm increases.[32] This is what I will do for the remainder of the paper.

3 A New Constraint

Here is a simple statement of a constraint against non-offset harm increases:

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

Like the old constraint against harm, we will likely have to complicate this by restricting its scope in various ways in order to allow unoffset harm increases to be permissible if required for self-defense, if they're consented

[32] We might instead decide to drop some larger assumptions, like the roughly commonsensical deontological framework we've been using. Consequentialists may beckon, bragging that their theories make short work of offsetting (like pretty much everything else). However, things are not so simple. For one thing, it's unclear whether consequentialists who want to go some way towards accommodating common sense intuitions about harm by fiddling with their axiology (and/or accepting an agent-relative, non-act, or non-maximizing version of consequentialism) are really in a better position to account for our judgments about offsetting. They may not need to revise any constraints, but they may have to make adjustments elsewhere. But even the consequentialists who keep their theories simple and are happy to pay for it by rejecting our judgments about various cases can't just say: a simple act that increases harm is permissible iff it brings about an outcome with the highest (expected) utility and the complex act of increasing harm and offsetting it is permissible iff it brings about an outcome with the highest (expected) utility. They can't just say this because there can easily be cases where the harm increase doesn't produce the best available outcome, but the harm increase together with the offsetting does. Adding to this a principle that says a permissible complex act cannot have impermissible parts, we get a contradiction. This is not a new problem for consequentialism—see Castañeda (1968) and Feldman (1986, Part I)—but nor is it one for which there is an easy solution. So it's not obvious that consequentialists are in a much better position to make sense of increasing harm and offsetting than the non-consequentialists are.

to, and so on. And we will need an account of how it interacts with other parts of morality and rationality. Presumably, for example, we want it to be permissibly infringed if infringing it is required for bringing about a great amount of good. Since such complications, as far as I can tell, are independent of whether we are talking about a constraint against harm or one against unoffset harm increases, I will not discuss them further here.

There are, however, some issues with the simple statement of the constraint that require discussion. One is that it runs into difficulties once we try to spell out just what it would be to offset some particular harm increase. We need an understanding of how harm decreases are linked to increases of the same harm, and it's not obvious how that should be done.

Take the zapping game plays of (DUU), (UDU), and (UUD). We want our constraint against increasing harm without offsetting to make such plays impermissible. How would it do so? In each case, it would say, there is some increase to the harm to Vicky that Perry did not and would not offset. We might think that this would mean that in each play, one of the U's is wrong and the other is permissible. It is pretty strange, though, to have to say this; what makes it true that it's *this* U rather than the other that was wrong? We could say, I suppose, that it's always the earliest unoffset harm that gets offset by the earliest equivalent harm decrease, so it's the second U in each case that was impermissible. But this seems arbitrary. We could just as easily say that it's the most recent unoffset harm that gets offset, in which case the first U's in each case would be the impermissible ones. And it would be pretty surprising if there were a brute fact which decides such matters one way or the other. It gets even worse once we have acts which increase and decrease harm by different amounts. Suppose we had something like (U[+0.5] U[+1] U[+0.5] D[-1]). Would there be two impermissible button presses here (the first two or last two U's), or just one (the middle)? And we get into even more trouble once we start to consider separate but simultaneous acts. If Perry presses two U buttons at the same time, but only does enough to offset the harm of one, 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 subset of the increases, we say that the effects of a decrease of some harm are distributed evenly over all the increases to that harm. So we'd say that in (UUD), for example, neither U is fully offset, so both are impermissible. And in order to explain why (UE), which only has one impermissible action, is just as bad as (UUDE), which has two, we'd 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 have a connection of degree of harm increase to stringency anyways, so there's nothing ad hoc about this.[33]

This idea about how to link decreases to increases of a harm is still underspecified, though, since there are different ways of spelling out what it is to 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 we had (U[+2] U[+1] D[-2]). Is an even distribution of the decrease to take away (the harm associated with an increase in) one point's worth of harm from each of the U's harm increases, or is it to take away two thirds of a point's worth from U[+2] and one third from U[+1]? Note that these accounts would differ on whether the second U was impermissible. The proportional construal is seems preferable, partly because it seems less likely to run into the kinds of arbitrariness worries we previously raised, and partly because it doesn't require any revision to deal with cases like like (U[+2]U[+1]D[-3]). The proportional construal correctly classifies this as a case of permissible offsetting—both of the harm increases are fully offset. But the amount construal, as it stands, seems to say that this involves a partially unoffset harm, the first U.

In any case, though, with the distributive (as opposed to direct linking) interpretation of offsetting, the proposed statement of the constraint is an attractive one. It's not clear to me, though, that it's the best way of formulating it. I'll now introduce another way to have a constraint against unoffset harm which seems to me roughly equally attractive. It applies to the whole contribution that one makes to a harm, rather than to the parts, so won't have to take on commitments about which of the parts are not fully offset and so wrong.

> Don't allow your net contribution to a harm be positive.

One thing about this proposal which might be thought to count against it is that it may allow for there to be impermissible complex actions (or

[33] Cf. the Aggravation Principle from Thomson (1990, p. 154).

collections of actions) of which there are no impermissible parts. I don't find such an implication troubling, however. Many cases of failing to do some required thing seem to have this property.

There is, though, a problem for either way of articulating the constraint which we have not yet discussed. The issue comes from the fact that harm increases to the same harm can contribute to different aspects of the harm, and harm decreases can take away from these different aspects. And it seems we should hold that harm decreases only offset harm increases to the same aspects of the same harm. If Perry were presented with buttons which would lead to increases or decreases to the same harm, but in different ways, it seems that he can't offset increases to one aspect of the harm with decreases to the other. He couldn't, for example, make a contribution that is net neutral in the relevant respect by increasing the duration but decreasing the intensity of the pain.[34] 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 trade-off. Thus I suggest we revise the constraint to be either

> Don't increase an aspect of harm if you have not and will not fully offset it,

or

> Don't allow your net contribution to any aspect of a harm be positive.

To fully understand the constraint, then, we'll need 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,[35] but I think we have enough of an intuitive grasp on them to have a working understanding

[34] This is not to say that decreasing the intensity at the cost of increasing the duration is always impermissible; it could be made permissible in the familiar ways of moral balancing, compensation, etc.

[35] Here is a first pass: an aspect of a harm h is a trope the existence of which is an 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; see Berker (forthcoming) for an argument against it) and gets us close to what we're after, this still doesn't work; it makes the event's being a half-second long an aspect of the harm to Vicky, whereas what we really want is that the event's *temporal length* to be an aspect. We might think we could easily generalize this first pass in the desired way by moving from a quantitative trope to the relevant non-quantitative,

of these reformulated constraints, at least as starting points for further theorizing.

Before moving on to do a bit of that theorizing, it is worth pausing here to emphasize that if we are serious about evaluating this new constraint, it will be crucial to work out the relevant metaphysical details of what harms and their aspects are and how each of these are individuated, as well as what exactly it is to increase an aspect harm and offset that increase. We need to do this both to see the degree to which the best version of the constraint fits our judgements about cases and to see if the metaphysical distinctions on which the constraint rests can be plausibly said to make a moral difference.[36] This is much as it is in the familiar debate about doing versus allowing harm,[37] and some of the same considerations can presumably be carried over to the debate about the new constraint, but much work remains to be done. At its current rather sketchy stage we simply do not have enough information to decide whether the new constraint should be, in the end, accepted. I will, however, leave the project of working out these details for another time.

4 Deriving the Old Constraint

I have argued that the kind of constraint against non-offset mere harm increase that would make (UE) but not (DUE) impermissible is not derivable from a standard constraint against harm, even taking into account various exceptions and ways for overriding such a constraint. And now I have given a couple ways of stating the new constraint that does make sense of

determinable trope of which the quantitative tropes are determinates. But we need to say more about how the relevant non-qualitative one is determined; being 0.5 seconds long is a determinate of the *having a length of time* trope, but it is also a determinate of the less general *having a length of time between 0 and 3 seconds* trope, and the more general *having a property* trope. How do we get from the trope of being a half-second long to the trope of having a length? This is a question I'll leave for elsewhere.

[36] Thanks to Shelly Kagan for making this point clear to me. He also raised some potentially problematic cases for distinguishing harm aspects which I hope to address elsewhere.

[37] See, e.g., Kagan (1989) and Bennett (1995) for attacks on the position that there is a morally relevant distinction between doing harm and allowing it and Kamm (2001), Kamm (2007), and Woollard (2015) for defenses. Woollard and Howard-Snyder (2016) give a survey of this and related topics.

increasing harm and offsetting.

The response to this could be to simply take on board the new constraint as an additional one, begrudgingly admitting that one cannot be derived from the other. Without some further story about the origins of these constraints, such a brute positing is unsatisfyingly disunified, especially given how similar a constraint against harm and a constraint against non-offset harm increase would be to one another and how much overlap they would have. And I think it is overly pessimistic, at this point, to punt to a foundational theory to do the unifying work for us, as there are still good prospects for deriving one constraint from the other. The derivation is just not in the direction we have already attempted. Instead of deriving the constraint against unoffset mere harm increases, we should go the other way, deriving a constraint against doing harm from the constraint against unoffset mere harm increases.

I won't work through the proposed derivation in detail, since that would require us to settle quite a few issues about how best to formulate each of the constraints, and about the nature of harm, events, causation, and so on. But the basics of the derivation can be given in an abstract enough way to apply more or less across the board.

The main idea behind the derivation doesn't involve anything complicated, just the observation that any harm of the normal sort will involve a net positive contribution to some aspect of a harm. 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 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 kind of moral work it does in the mere harm increase cases, but rather because it's impossible to fully offset your increase to harm but still have done harm. If the increase to harm had been fully offset, there would have been no harm at all.[38]

A harm in the familiar sense, then, is also a harm increase that is not fully offset. Thus any violation of the old constraint against harm will also

[38] 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 section 2.5. 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.

be a violation of the new constraint against unoffset harm increases, so there's no reason to have the old constraint in addition to the new one. The new constraint does all the work the old one did, plus, as we've seen, some work that it can't do. We should replace the traditional constraint against harm with a constraint against unoffset harm increases.

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 (forthcoming). "The Unity of Grounding".

Bradley, Ben (2012). "Doing Away With Harm". In: *Philosophy and Phenomenological Research* 85.2, pp. 390–412.

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

Castañeda, Hector-Neri (1968). "A Problem for Utilitarianism". In: *Analysis* 28, pp. 141–142.

Feldman, Fred (1986). *Doing the Best We Can: An Essay in Informal Deontic Logic*. D. Reidel Publishing Company.

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

— (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.

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.

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. (2001). *Morality, Mortality Volume II: Rights, Duties, and Status*. Oxford Univerisity Press.

— (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.

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

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

Parfit, Derek (1984). *Reasons and Persons*. Oxford University Press.

Paul, L. A. (2002). "Logical parts". In: *Noûs* 36.4, pp. 578–596.

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

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/.