Doing Harm *While* Allowing It: Towards a New Framework for the Doctrine of Doing and Allowing

David Turon - dturon@umass.edu

Pre-print Version - Forthcoming in *Philosophers' Imprint*
(https://doi.org/10.3998/phimp.3370)

Abstract: Doing harm and allowing harm are not contraries. In fact, in most instances in which one does harm, one also thereby allows harm. I argue that, on its own, doing harm (without allowing it) carries little or no moral weight -- surprisingly, less moral weight than merely allowing harm. On the other hand, doing harm while allowing it is significantly harder to justify than merely allowing harm. I argue that ordinary harm-doing is especially hard to justify because it involves a convergence of doing and allowing harm. I show that this is not simply the result of an additive effect, but rather, there is a moral synergy between doing harm and allowing the same harm to the same victim. I conclude by offering two possible explanations for the synergy, helping us to better clarify and recast the Doctrine of Doing and Allowing as well as to determine its justification.

1. Introduction

Generally speaking, it is seriously wrong to do harm to others. It is also often seriously wrong to allow harm to others. Some hold nonetheless that doing and allowing harm are morally inequivalent; they endorse the Doctrine of Doing and Allowing (DDA), the view that it is harder to justify doing harm than *merely*

allowing harm, all else being equal.[1] For example, it seems wrong to deflect a lethal threat onto an innocent in order to save oneself but permissible to allow a lethal threat to reach an innocent in order to save oneself, and the DDA naturally explains why. But others deny that there is any morally significant difference, arguing that when all else is equalized, doing harm is no worse (nor harder to justify) than allowing harm.[2]

This way of putting the issue pits doing harm against allowing harm. But why should we pit them against each other? Does one do some harm only if one does not allow it, and one allow harm only if one does not do it? Are doing and allowing mutually exclusive? I do not think so.[3] And recognizing the ways in which doing and allowing harm can come together and come apart, as I shall argue, reveals a surprising amplificatory relationship between doing and allowing harm. Explaining this amplificatory relationship may clue us in to some of the reasons why most ordinary harm-doing (which involves *both* doing and allowing harm) is harder to justify than merely allowing harm.

---

[1] I draw this way of characterizing the DDA from Fiona Woollard (2015). For a few examples of non-consequentialists who have defended the Doctrine of Doing and Allowing, see Woollard (2015), Foot (1967, 1984, 1985), and Quinn (1989). See also Liu (2012), who offers a consequentialist rationale for the doing/allowing distinction.

[2] See, e.g., Bennett (1995) and Rachels (1975).

[3] Some authors writing on the DDA (e.g., Woollard, 2015) have been clear to avoid this commitment, focusing on the contrast between doing harm and *merely* allowing harm, where the latter is explicitly defined in a way that makes it exclusive with doing harm.

Here's a roadmap for the paper. In §2, I argue that when it comes to doing and allowing harm, we should distinguish between (i) *merely* allowing harm, (ii) *merely* doing harm, and (iii) doing harm *while* allowing it. Surprisingly, on its own, doing harm seems to carry little or no moral weight; in fact, it seems *easier* to justify *merely* doing harm than *merely* allowing harm. But doing harm *while* allowing it (what happens in most cases we intuitively classify as *doing harm*) seems harder to justify than merely allowing it, as the DDA proclaims.

In §3, I argue that, though there may be independent moral weight associated both with doing harm and allowing harm, the moral weight of one or the other (or both) is *amplified* when the two converge; there is a moral synergy between doing and allowing harm. In particular, I argue that it takes much more to justify doing harm to someone while allowing them to suffer that harm than it does to *merely* do an equivalent harm to one while *merely* allowing an equivalent harm to another. This gives rise to a puzzle: why is it so much harder to justify doing some harm while allowing *that* harm than merely doing some harm plus merely allowing some distinct but equivalent harm?

In §4, I sketch two distinct but compatible ways of resolving this puzzle. First (in §4.1), I explore the possibility that introducing a threat of harm generates a special duty not to allow that harm, thereby making the duty not to allow harm more stringent. Second (in §4.2), I explore the possibility that the rights-infringements involved in doing harm constitute graver wrongs the more harmful they are for the victim (in line with Judith Jarvis Thomson's (1990) Aggravation Principle). I conclude in §5 by suggesting how we could recast the DDA to better reflect the relationship between doing and allowing harm.

2. Doing *Versus* Allowing Harm?

Some have distinguished standard instances of harm-doing from cases of allowing oneself to do harm.[4] For example, suppose someone poisons a victim's drink but has the opportunity later to warn the victim before she takes a sip. In refraining from warning the victim, the poisoner allows himself to have done harm to the victim. Thus, he both does and allows harm to the victim. But rather than taking instances in which an agent both does and allows a harm to be a special kind of case, we should recognize that *most* (but *not* all) instances of what is ordinarily regarded as *doing* harm are also instances in which the harm-doer *allows* the harm in question. To see why, let's first get clearer on what it is to allow harm.

The rough, intuitive idea is this: you allow someone to be harmed just in case they suffer harm and you could have done something such that, if you had done it, they wouldn't have suffered that harm.

More precisely, an agent allows harm of magnitude *m* to a subject if and only if (i) that subject suffers some token harm *H* of magnitude *m*, and (ii) the agent had an option available to her such that, had she taken it, the subject would neither have suffered *H* nor have suffered any harms of total magnitude *m* (or greater) instead of *H* that he didn't actually suffer.[5]

---

[4] Judith Thomson pointed this out in her (1985: 1400). There has been more recent controversy about the moral significance of allowing oneself to do harm (see, e.g., Hanna, 2014; Unruh, 2021).

[5] I am indebted to an anonymous referee for pushing me to be clearer about 'allowing harm' here. In particular, we should take care to acknowledge the distinction between failing to avert a *token* harm, failing to avert harm of a certain *kind*, and failing to avert a *total magnitude* of harm. My

Here's an example. Suppose you trip on a log and break your leg, suffering some associated harms. Bob could easily have removed the log, but he didn't. If Bob had removed the log, you wouldn't have suffered the harms associated with breaking your leg (nor would you have suffered any new harms of equal or greater magnitude that you didn't actually suffer). Thus, Bob allowed you to suffer harm of a certain magnitude, namely, the total magnitude of the harms associated with breaking your leg.

---

proposed definition falls into the third category, though it mentions particular harms in the *definiens*. This proposal will ultimately deliver equivalent results regardless of how modally fragile we take token harms to be, that is, how much a harm's features can vary and still count as *the same harm*.

Here's an example to illustrate why. Suppose you have two options: option (a) on which Ann will get a bruise, and option (b) on which she will get a bruise at a slightly different place on her body that causes or constitutes harm of exactly the same magnitude as that which she suffers on option (a). All else is equal. Whether we classify the harm arising from the bruise when taking option (a) as *the same harm* as that arising when taking option (b), you won't have allowed her to suffer any magnitude of harm. First, more simply, if Ann counts as suffering the same harms on options (a) and (b), then she would have suffered all the same harms no matter what you did, and so you won't have allowed her to suffer any magnitude of harm. If she doesn't count as suffering the same harms on options (a) and (b), then on option (b), Ann will suffer a different harm that is of the same magnitude as the harm she suffers on option (a). So again, by my definition, you would not count as allowing her to suffer any magnitude of harm no matter which option you take.

With the notion of allowing harm so clarified, we can now see that there is a simple reason why when someone does harm of some magnitude, they also usually allow harm of that magnitude. Consider the following case:

> *Button*: Ann presses a button. Once this button is pressed, it sends a current that will shock Bob a few seconds later. Ann is unable to do anything to stop the shock from reaching Bob once she presses the button.

It is clear that Ann *does* harm of a certain magnitude to Bob by pressing the button. But she also simultaneously *allows* Bob to suffer that magnitude of harm, for at the moment she pressed the button, she had the ability to do something else instead (e.g., sit still) such that, had she done it, Bob would not have suffered a harmful shock soon after. However, in *Button*, Ann loses the option to avert the harm that Bob suffers the instant she initiates the threat.

Here is the upshot: even if one cannot undo a threatening sequence at any moment after it is set in motion, it can still be true that one allowed someone to suffer some magnitude of harm from the very harms caused by the threatening sequence. This is usually because one could have refrained from launching the threatening sequence in the first place. For this reason, I submit, most instances of doing harm are also instances of allowing harm, not simply in virtue of the harm-doer's having an ability to cancel a threat *after* it is launched, but simply in virtue of the fact that, most times when someone does harm to a victim, (i) she had the ability not to launch the threat of harm in the first place, and (ii) had she

not launched the threat in the first place, the victim would not have suffered that harm.

Why, then, might we have been inclined to describe Ann as doing harm *rather than* allowing harm to Bob in cases like *Button*? We might think this is because the meanings of ⌜doing ϕ⌝ and ⌜allowing ϕ⌝ are such that one entails the negation of the other. But if that were so, I could simply say that the morally relevant notions that I refer to by 'doing harm' and 'allowing harm' are distinct from the referents of those expressions in ordinary English. Nonetheless, I think that 'doing' does not preclude 'allowing' (nor vice versa) even in the ordinary sense. I would submit that describing someone as allowing harm merely conversationally implicates that she only allowed harm without doing it. Again, most of the time when we do harm, we also thereby allow it. So normally, it goes without saying that if one does harm, one allows it (it typically goes without saying that, if you did harm to someone, you could have done something else such that, had you done it, the victim would not have suffered harm). Given this, describing someone as allowing harm conversationally implies that she did not do harm, as omitting this further piece of information would run afoul of the Gricean Maxim of Quantity.[6] This would explain why we might be inclined not to describe Ann as allowing harm in *Button*.

To support this diagnosis, observe that we can cancel the relevant implicature that flows from describing someone as allowing an outcome, using our ordinary sense of 'allow'.

[6] See, e.g., Grice (1975).

(1) Did you allow the cat to go outside?

Yes. In fact, I put the cat outside myself.

(2) Did you allow the bomb to go off?

Yes. In fact, I detonated it myself.

These exchanges check out to my ears.

Note again that it would not be a problem if the notion of *allowing harm* as I've defined it above does not fully comport with ordinary usage of the English term 'allow'. For the notion of *allowing harm* as I've defined it, as I shall argue below, is a morally significant notion that has a synergistic relationship with *doing harm*, and that remains so even if it does not completely match up with the ordinary usage of the English expression. But in any case, I suggest that the expression ⌜allowing ϕ⌝, as it is ordinarily used in English, *does* fit with my proposed definition – that is, it does not preclude ⌜doing ϕ⌝ or ⌜ϕ-ing⌝, and the temptation to think otherwise is, I suggest, best explained by conversational implicature, not by the semantics of 'do' and 'allow'.

To define *doing* harm *while allowing it*, we can now simply take our earlier definition of allowing harm and add the condition that the agent *does* the harm of the relevant magnitude that was actually suffered.

A does harm of magnitude *m* to *V* while allowing it $=_{\text{def}}$ (i) *V* suffers a token harm *H* of magnitude *m*, (ii) *A does* this harm to *V*, and (iii) when she did this harm, *A* had an option such that had she taken it, *V* would neither have suffered *H* nor any harms of *m* or greater total magnitude that she didn't actually suffer.

So far, I have argued so far that, when we do harm to someone of some magnitude, we *usually* also allow them to suffer that magnitude of harm. In §2.2 below, however, I argue that this is not *always* the case.

2.2. Merely Doing Harm

Let's define *merely doing* harm as doing harm of some magnitude without doing so while allowing it. Is it possible to *merely* do harm? What would *merely* doing harm be like?

Given the above definition of doing harm of some magnitude *while* allowing it, to merely do harm is to do some harm while lacking an option to do anything else such that that harm (or an equivalent harm taking its place) wouldn't have occurred.[7] Thus, we *merely* do harm when we do harm to someone who is, relative to what is in our control, doomed to suffer that harm (or equivalent harm) no matter what we choose. As I shall argue now, *merely* doing harm is relatively *easy* to justify – easier, in fact, than *merely* allowing harm. Thus, perhaps surprisingly, doing harm carries little or no moral weight on its own. Here's an illustrative example of *merely* doing harm:

[7] If we look to the above definition of *doing harm while allowing it* in particular, we *merely* allow harm when we satisfy all but the second condition, and we *merely* do harm when we satisfy all but the third condition.

Assured Crushing: A machine will release a boulder on Tara at time *t*, lethally crushing her, unless a certain button is pressed before *t*. Bob can stop this machine from releasing the boulder only by pressing this button. However, pressing this button will trigger a different machine to release the boulder on Tara at *t*, lethally crushing her in just the same manner. So whatever Bob does, the boulder will fall on Tara at *t*, causing her to be lethally crushed in the same manner either way.

If Bob presses the button himself, *he* will have killed Tara. Why? *He* will have initiated the sequence that culminates in Tara's death. But Bob does not allow Tara to suffer harm. Why not? First, to simplify, let's say that Tara suffers just one relevant harm – namely, being crushed by the boulder. Bob lacked the ability to do something such that, had he done it, Tara would not have been crushed by the boulder. Thus, there is no relevant harm to Tara such that, had Bob taken a different option, Tara would not have suffered it or an equivalent harm in its place; with respect to Bob's options, Tara was doomed to suffer the same magnitude of harm from being crushed by the boulder.

Now, I don't think it ultimately matters whether Tara is doomed to suffer the *very same* token harm. What seems to matter is whether Tara is doomed to suffer *either* the harm she actually suffers *or an equivalent* harm that she didn't actually suffer. To motivate that suggestion, first consider how we would morally evaluate Bob's conduct in *Assured Crushing*. Suppose Bob presses the button. In that case, Bob does harm to Tara at *t* but does not allow harm to Tara. Does he act wrongly? The answer is not immediately obvious. It does seem that we would

want *some* reason from Bob as to why he pressed the button. That may suggest that Bob needs *some* justification to press the button. But on the other hand, the considerations that would permit Bob to push the button in *Assured Crushing* could be much weaker than the considerations that would be needed if the boulder were *not* otherwise going to be dropped; Bob is morally permitted to push the button in *Assured Crushing* if doing so is necessary to save someone else's life, but the prospect of saving a life would not be enough to justify pushing the button if the boulder was not otherwise going to be dropped on Tara. (That would be to kill one non-doomed innocent in order to save another, which I take to be morally wrong in general.) In *Assured Crushing*, though, it strikes me that Bob would be permitted to push the button even to prevent a much lesser harm for someone else, e.g., a broken leg.

Now, notice that these judgments seem to hold even if Bob's pressing the button were to cause Tara to suffer a different harm, so long as we are sure the total magnitude of harm to Tara remains the same. Suppose pressing the button caused Tara to be killed by a trolley at *t* instead of a boulder, and suppose that being killed by the boulder or the trolley would produce the same total magnitude of harm for Tara (suppose she is unconscious and will feel no difference either way).[8] This does not seem to make a moral difference; it seems that Bob would still be permitted to press the button in order to save someone else's life, or even

[8] If she were conscious, this might introduce unwanted noise into our judgments, as we might have a recalcitrant doubt that being killed one way really is just as harmful as being killed in the other.

to prevent a broken leg, even though in doing so, he does lethal harm to Tara while having the option to do something else such that Tara would not have suffered *that* very same harm. Thus, the morally relevant notion of allowing harm seems to be tied to the total magnitude of harm that someone suffers, either from the very harms that she actually suffers *or* from equivalently severe harms she didn't actually suffer but would have suffered instead.[9] In the trolley-versus-boulder variation on *Assured Crushing*, Tara is doomed to suffer the same magnitude of harm no matter what Bob does; Bob's pressing the button simply replaces one harm with an equally severe harm and likewise seems relatively easy to justify.[10]

---

[9] An anonymous referee asks: what if, by doing some harm to Tara, one thereby prevents Tara from suffering an equivalent harm in the distant future, which Tara was otherwise doomed to suffer? Does it matter whether the harm that one replaces happens at or around the same time? For present purposes, I'll remain neutral on this and explain how my proposal could be tweaked to accommodate our moral judgments whichever way we settle the question. We can require that, for some harms to count as happening *instead* of *H*, they have to happen at or around the same time as *H* would have happened. All examples used in this paper are consistent with a narrower, time-sensitive notion of *allowing harm* that builds this more stringent requirement into it; nothing I shall argue for here rules out, or requires, this narrower definition of 'allowing harm'.

[10] If one were to insist that it *does* matter whether the harm Bob causes Tara by pressing the button is *the very same harm* she would have suffered otherwise – so that pressing the button is easier to justify in the original version of *Assured Crushing* than in the trolley-versus-boulder variant – I could still make the same points about the relationship between doing and allowing harm; I would simply need to contrive the examples in the next section so that the relevantly

I would submit, then, that even if there is some moral pressure against doing harm without allowing it, this moral pressure is relatively easy to override. In fact, it is weaker than the moral pressure against *merely* allowing harm. For compare *Assured Crushing* with the following case: suppose the machine were set up so that pressing the button would save Tara's life (by preventing the boulder from killing her without releasing any other threats). Bob would need more justification to justify *refraining* from pushing this button than he would need to justify pushing the button in *Assured Crushing*. The prospect of preventing a broken leg would *not* be enough to justify Bob's refraining from pressing this button if it would save Tara's life. For when we must choose between saving a life and preventing a broken leg, I take it that we must save the life. But again, it would be permissible for Bob to press the button in *Assured Crushing* in order to prevent someone from breaking a leg. So doing harm without allowing it – *merely* doing harm – is actually *easier* to justify than *merely* allowing harm! This claim may seem at odds with the DDA, but it need not be. It is perfectly coherent to accept the DDA (at least in a suitably clarified form) while also accepting that, on its own, doing harm carries *less* moral weight than allowing harm. We shall see how this can be so in §§3-4.

Let's sum up what we've established so far. Most times when an agent does harm to someone, she does a magnitude of harm while allowing it. However, it is possible for an agent to do harm without allowing any harm

---

doomed individuals are doomed to suffer the same exact harm no matter what. I have opted not to do so, simply because this would further complicate already complicated examples.

(though this is uncommon) and possible for her to allow harm without doing it. Thus, doing and allowing are two separable ways that we can relate to harms, and relating in one way to a harm does not preclude our relating in the other way to the same harms. There may be moral pressure against relating in either way to harms. However, the moral pressure against merely doing harm, i.e., doing harm without allowing it, is weaker (*viz*. easier to override by other considerations) than the moral pressure against merely allowing harm, i.e., allowing harm without doing it.

3. A Moral Synergy Between Doing and Allowing Harm

It may be tempting to conclude that what really matters in the above-considered cases is the difference that we make for the worse for those who suffer harm. That is, we might think that the moral status of an act is determined by the comparison between how well everyone would fare if the act were performed and how well everyone would fare had some alternative act been performed instead. Or perhaps, in a less consequentialist vein, we might think that the justification required to permissibly do harm is determined by how poorly off the harm-doing makes the harmed subject in comparison with how that subject would have fared had the act not been performed. Because things go just as well for everyone whether or not Bob pushes the button in *Assured Crushing*, either view could explain the judgment that Bob is permitted to push the button even though doing so kills Tara.

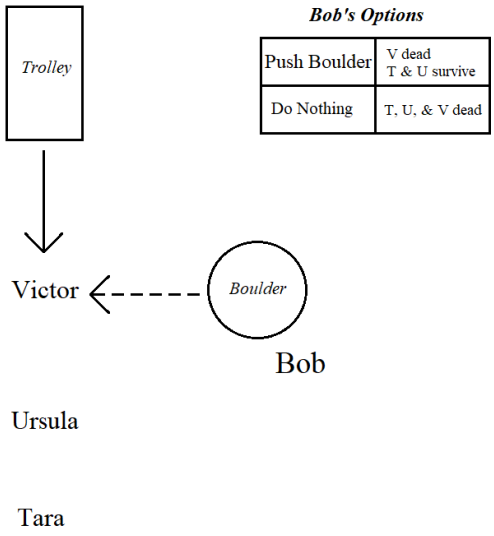
However, I do not think either view is correct. Reflection on further cases suggests instead that there is an amplificatory relationship between doing harm

and allowing harm, but only when we do harm while allowing it to the same subject.

As a baseline, start with the following case in which one *merely* does harm:

Three Trapped (One Doomed): A trolley is heading for Tara, Ursula, and Victor, who lie unconscious on the tracks. If Bob does nothing, all three will be killed by the trolley. No matter what Bob does, Victor will be crushed to death before he wakes up; Bob cannot wake up any of the victims. However, Bob can roll a boulder over Victor, crushing him, but preventing Tara and Ursula from being crushed by the trolley.

Figure 1: Three Trapped (One Doomed)



As with the earlier variation on *Assured Crushing*, I think Bob is permitted to crush Victor in *Three Trapped (One Doomed)*. Victor is doomed to be

crushed to death before he wakes up, no matter what Bob does. If Bob does nothing, Tara, Ursula, and Victor will all die. If he rolls the boulder, *he* kills Victor, but at least Tara and Ursula will be saved. Bob may need a reason to push the boulder, but he has it in this case: the prospect of saving two lives.

Now, let's contrast *Three Trapped (One Doomed)* with a few other cases. First, consider Philippa Foot's *Rescue II* case,[11] which I have adapted below in a slightly modified form:

> *Rescue II*: Five people are in imminent, lethal danger. To save them, you must drive to them immediately and administer life-saving medicine to each of them. Unfortunately, to reach them before they die, you would have to traverse a narrow road, on which one person is trapped. Thus, you can either do nothing, letting the five die, or you can proceed to rescue the five, running over and killing the one person trapped on the road.

Foot claims that it is impermissible to run over the one person trapped in the road, even to rescue the five others. I agree with Foot. This judgment stands in striking contrast to *Three Trapped (One Doomed)*. In *Rescue II*, it seems wrong to kill one to save five. In *Three Trapped (One Doomed)*, it seems permissible to kill one to save two. What explains the difference? Perhaps, one is tempted to think, it is the fact that the one trapped in the road in *Rescue II* would

[11] Foot (1984: 179).

not be under any threat of imminent death if we abstained from saving the five. In *Three Trapped (One Doomed)* by contrast, Victor is under threat of imminent death no matter what we do.

This explanation surely seems to be on the right track, but I would submit that it is not satisfactory by itself. For consider the following variant of Foot's *Rescue II* case:

> *Rescue III*: Things are as they were in *Rescue II*, except the one person trapped in the road will imminently die unless you stop and provide life-saving medicine to her instead of providing that life-saving medicine to the five further down the path.

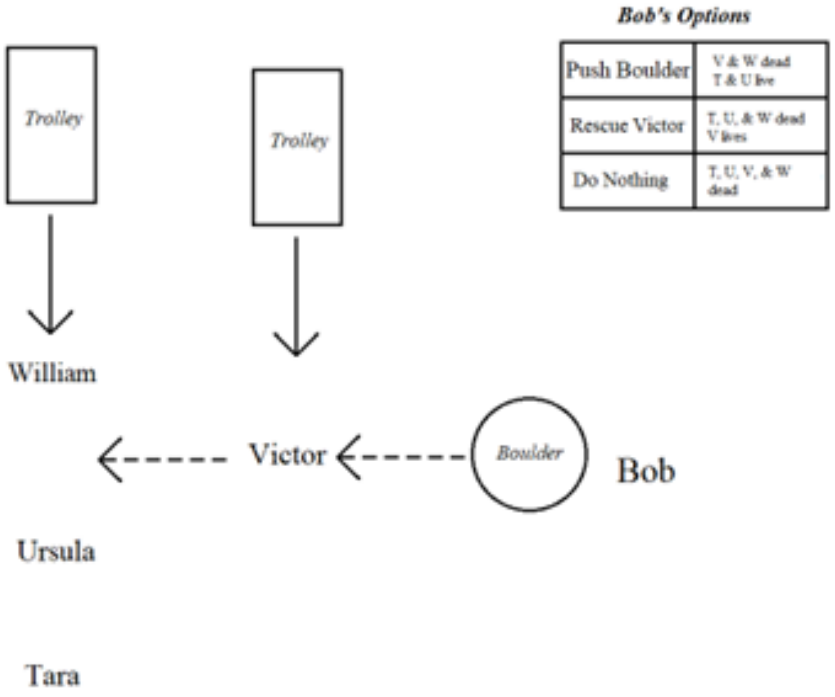
In *Rescue III*, you have three options: (i) do nothing, in which case all six people die, (ii) save the one person trapped in the road, letting the five further down the road die, or (iii) run over the person in the road, killing her, and save the five people further down the road. Here, it seems to me that option (ii) is the only permissible option. It does not matter that the person trapped in the road requires your aid to survive; it still seems wrong to run her over, just as it would be to do so in the original *Rescue II* case. What seems to explain the difference between *Three Trapped (One Doomed)* and *Rescue III* is the fact that you are *able* to do something that would result in the person in the road's survival in *Rescue III*, whereas you are not able to do anything that would result in Victor's survival in *Three Trapped (One Doomed)*. This suggests that doing harm

becomes harder to justify not only when abstaining would have resulted in less harm for the victim, but when we have some *option* that would have resulted in less harm for the victim (that is, when we simultaneously *allow* the victim to suffer harm).

Given this judgment, let's consider a variant of *Rescue II/Rescue III* that is closer to *Three Trapped (One Doomed)*:

> *Convergence*: Two trolleys are out of control, heading toward people who, through no fault of their own, lie temporarily unconscious on the tracks ahead of them. One is heading for Tara, Ursula, and William, and the other toward Victor. If Bob does nothing, all four will be killed by the trolleys. Bob can roll a boulder to the tracks behind William, thereby preventing Tara and Ursula from being killed by the trolley. However, if Bob does this, the boulder will crush Victor, killing him, en route to saving Tara and Ursula. Bob can also save Victor from the trolley by shouting loudly, thereby waking him (but not the others) and causing him to move himself off the tracks. However, if Bob does this, he will be unable to move the boulder in time to save Tara and Ursula from the trolley. Tara and Ursula are too far away for Bob to wake them.

Figure 2: Convergence



(Note that the reason for including William in the case, for whom Bob can do nothing, will become apparent when we compare *Convergence* with *Divergence* below.)

I would submit that Bob may not crush Victor with the boulder in *Convergence*, just as we would not be permitted to run over the person trapped in the road in *Rescue III*. Given that Victor is not doomed in this case (because Bob can easily save him by shouting), Bob is not permitted to kill Victor to save Tara and Ursula. Harming one to save two from equivalent harms is permissible, as in *Three Trapped (One Doomed)*, if the one is doomed to suffer the same total amount of harm at the same time no matter what one chooses. But killing one to save two is not permissible if the agent can somehow avert the one's death. In that case, in killing the one, the agent *does* allow the one to suffer harm. Bob is

not permitted to kill Victor in *Convergence* because he could have saved Victor, and thus, in killing him, he would kill Victor *while* allowing him to die. Among the remaining options, i.e. easily rescuing Victor or doing nothing, I take it that the only permissible option is to rescue Victor.

This has a number of upshots. By pushing the boulder over Victor in *Convergence*, Bob need not make any difference for the worse for Victor; Victor would have suffered the same total amount of harm had Bob done nothing. And even if we suppose that Bob is so disposed that, had he not pushed the boulder over Victor, he would instead have done nothing (and thus, had Bob not pushed the boulder, Victor would have suffered the same total amount of harm), Bob would still not be permitted to push the boulder. Thus, what appears to matter when an agent does harm is not necessarily the difference that an agent makes in comparison with what would have happened had the agent done nothing, or even in comparison with what would have happened had the agent not acted as she did. What matters in determining the wrongness of doing harm is whether an alternative was *available* for the harm-doer such that, had the harm-doer taken that alternative instead, the harmed subject would have suffered less total harm.[12]

[12] One caveat here: it may matter whether the available alternative in question is morally permissible. For example, if Bob could only save Victor by means of killing an innocent bystander, it may then be permissible for Bob to roll the boulder over Victor after all; perhaps the moral weight of Bob's allowing Victor to die depends on whether Bob had an ability to *permissibly* do something that would have resulted in Victor not dying. If he doesn't, then perhaps morally speaking, it is as if Bob doesn't count as allowing Victor to die. Thanks to Peter A. Graham for bringing this point to my attention.

When our choice is between allowing harm to one and allowing harm to another, all else equal, we are permitted to choose either way. But *Convergence* suggests that when our choice is between doing *while* allowing harm to one and merely allowing harm to two others (all else equal), we are forbidden from doing *while* allowing harm to the one and must instead allow the others to be harmed. This remains so even if by doing while allowing harm to one, the agent leaves the one just as badly off as she would have been had the agent done nothing (or in comparison with how well off she would have been had the agent not performed the action she performed).

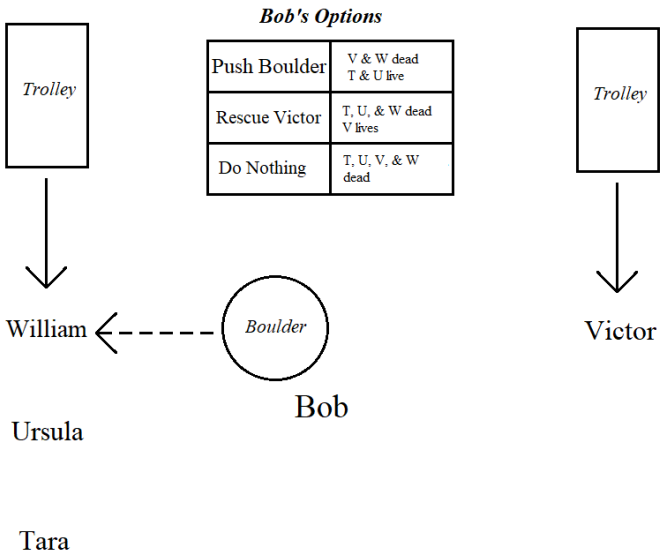
But this leaves open a further question: does the stringent prohibition against doing *while* allowing harm to a subject simply stem from the fact that, when we do harm while allowing harm to one, we perform an act that is *prima facie* wrong in two separate ways, where each way independently adds to the threshold of justification? If, supposing for simplicity's sake, stealing an apple needs to produce X amount of good to be justified, and knocking someone down needs to produce Y amount of good to be justified, then stealing an apple by pushing someone down needs to produce X + Y amount of good to be justified.[13] Is the stringent prohibition against doing *while* allowing harm to a subject simply additive like this?

I submit that it is not. If it were, then we should expect that Bob would also be forbidden from crushing William in the following case.

[13] Cf. Kagan, 1988.

Divergence: A trolley is heading for Tara, Ursula, and William, who lie unconscious on the tracks. Off to the side, another trolley is heading for Victor. If Bob does nothing, all four will be killed by trolleys. No matter what Bob does, William will be crushed to death before he wakes up. However, Bob can roll a boulder over William, crushing him, but preventing Tara and Ursula from being crushed by the trolley. Alternatively, Bob can run to Victor and wake him by shouting loudly, but then Tara, Ursula, and William will all be killed by the other trolley.

Figure 3: Divergence



I submit that it is permissible for Bob to push the boulder onto William (precluding him from saving Victor) in order to save Tara and Ursula. *Divergence* is just like *Three Trapped (One Doomed)*, except that refraining from pushing the boulder would give Bob an option to save one other person off to the side (Victor in this case). Since William is doomed to suffer the same magnitude

of harm regardless of how Bob acts, it seems that Bob is permitted to push the boulder in order to save a greater total number of lives, just like he is in *Three Trapped (One Doomed)*.[14]

Notice that just like in *Convergence*, by pushing the boulder in *Divergence*, Bob does lethal harm to one and allows lethal harm to another. Look at the overall consequence profile of Bob's options in terms of who lives and who dies, as well as the number of people Bob kills and allows to die. The two cases are exactly the same in these respects. In *Convergence* and *Divergence*, Bob can either (i) do nothing, (ii) rescue Victor, or (iii) push the boulder. Doing nothing results in Tara, Ursula, Victor, and William all dying (Bob kills no one but allows Victor, Tara, and Ursula to die -- because he could have saved them -- though he does not allow William to die, because he lacks the ability to save William). Rescuing Victor results in Victor's survival but the deaths of everyone else (Bob kills no one but allows Tara and Ursula to die, though he again does not allow William to die because he lacks the ability to save William). Pushing the boulder results in Tara and Ursula surviving but Victor and William dying (Bob kills one and allows one to die). What's the difference, then? In *Divergence*, Bob kills a different person than the one he allows to die; he kills William without allowing William to die, and he allows Victor to die without killing Victor. In contrast, in *Convergence*, he both kills *Victor* and allows *Victor* to die.

[14] Note that I do not assume that Bob is morally *required* to crush William; all that matters for present purposes is that he is permitted to do so.

This, I submit, is why Bob is permitted to push the boulder in *Divergence* but not in *Convergence*; doing and allowing harm is harder to justify when we do the very harms in virtue of which we allow a magnitude of harm to be suffered. Even though Bob does lethal harm to one person and allows lethal harm to one person in *Divergence*, the fact that the one Bob kills is a different person from the one he allows to die seems to make his action more easily justified, *viz.* justified by the prospect of saving two other people's lives.

There appears, then, to be a synergistic relationship between doing and allowing harm.[15] All else equal, when we do harm to someone while allowing them to suffer the harm we did to them, this is harder to justify than when we merely do harm to one and merely allow harm to another. Either doing harm to someone amplifies the justification needed to allow them to suffer harm, allowing harm to someone amplifies the justification needed to do harm to them, or both amplify each other.

So there is a moral synergy between doing and allowing harm. But we should also wonder *why* there is such a synergy. In the following section, I sketch two possible explanations.

[15] Or they form a kind of Moorean organic unity. See Moore (1903).

4. Two Routes to Amplification

4.1. Amplifying the Moral Weight of Allowing Harm

Let's start with an explanation as to why doing harm could make it harder to justify allowing total harm to which it contributes.

In general, some justification is needed in order to permissibly allow harms. *More* justification is needed to permissibly allow harms for which one has a special duty not to allow. For example, more justification is needed to allow one's own child to suffer harms than to allow another's child to suffer the same harms. If we must choose between saving our own child from harms and saving another's child from similar harms, we must save our own child; we owe it to her to save her. This is perhaps because, as caretakers, we take responsibility for those under our care, giving us a special obligation not to allow those under our care to suffer harm.[16]

Perhaps doing harm is another way to acquire a special obligation not to allow harm. By creating a threat of harm, an agent thereby becomes morally responsible for what that threat goes on to do, including the harms it causes. Those who launch threats of harm to others thereby acquire special obligations not to allow them to suffer total harm from such threats, over and above their general duty not to allow others to be harmed. This idea would also account for the intuitive judgment discussed, e.g., in Hanna (2014) that, after a threat has

[16] Of course, if those in our care suffer harm in a way that we are powerless to stop, we have not wronged them, because we have not allowed them to suffer harm; if we never have the ability to stop something from happening, it is false that we allowed it to happen.

been launched, and one faces a choice between saving one's future victim from this threat or saving two unrelated others from equivalent threats, we should save the one from our own threat.[17]

When Bob allows Victor to die in *Divergence*, he allows Victor to suffer harm that he has no special duty to prevent. This is *prima facie* wrong (because, in general, there is a *prima facie* duty not to allow others to suffer harm), but *ultima facie* justified by the fact that the only way to save Victor requires Bob to allow both Tara and Ursula to die instead. In *Divergence*, although Bob introduces a threat that culminates in William's death and thereby acquires a special duty not to allow William to suffer harm from this threat, Bob does *not* allow William to suffer harm. This is because he lacks the ability to avert it; he could never have done anything else such that, had he done it, William wouldn't have suffered the harm that he actually suffered (or any harm of equal magnitude). Since Bob does not allow William to die, he *a fortiori* does not flout any duty not to allow William to suffer harm.

In *Convergence*, by contrast, Bob is not permitted to allow Victor to die. His option to push the boulder on Victor is morally unacceptable because it *does* flout a duty not to allow Victor to die. As with William in *Divergence*, Bob's pushing the boulder introduces a threat to Victor's life and thereby gives Bob a special obligation not to allow Victor to suffer total harm from that threat. But

[17] Cf. Unruh (2021) for a similar proposal on which doing harm (or rather, launching a threat of harm) amplifies the moral pressure not to allow harm by giving rise to a special duty not to allow that harm.

unlike *Divergence*, in which Bob does not allow William to die, in *Convergence* Bob has the ability to prevent Victor's death, and thus Bob does flout a special duty not to allow Victor to die when he kills him while also allowing him to die. Although Bob could save two others by killing Victor while also allowing him to die, this is not enough to justify Bob's allowing Victor to die. By launching the lethal threat toward Victor, Bob generated a special obligation in himself not to allow Victor to die, but he lacks such special obligations with regard to Tara and Ursula. If the only way Bob can satisfy this special obligation to Victor is not to roll the boulder toward Victor in the first place, then he is not permitted to do so.

To repeat, on this explanation, *allowing* harm does the direct moral work, and *doing* harm is an aggravating factor; it alters the moral status of allowing harm (in particular, increasing the threshold for its justification). Note, however, that for this story to work on its own in explaining the moral synergy between doing and allowing harm, we may need to add that the moral weight of special duties to prevent harm comes in degrees. For we might hold that we are not permitted to kill an innocent bystander to save our own child's life, yet we do want to say that we have special duties to prevent harm to children under our care.[18] If the wrongness of killing the bystander is explained by the fact that, in doing so, we allow a harm for which we have a special duty not to allow, then why

[18] In a similar vein, notice that if someone wrongfully pushes an innocent in front of a trolley and subsequently must choose between saving this innocent or their own child, we might think that they must save the innocent they put in harm's way. Thanks to Peter A. Graham for suggesting this case.

wouldn't killing the bystander while allowing her to die be morally *equivalent* to allowing our child to die? In either case, one allows a death for which one had a special duty to prevent. If the strength of special duties comes in degrees, however, we can explain why the duty not to kill the bystander while allowing her to die wins out by holding that doing harm gives us a *stronger* special duty not to allow said harm than the special duties that caregivers have not to allow harm to come to those under their care.

Of course, we would want to know why this is so, and for now, I leave it open how such a story would go and whether it could be justified on independent grounds. In the following section, though, I offer an alternative that might either supplement or replace the story just considered.

4.2. Amplifying the Moral Weight of Doing Harm

In this section, I explore an alternative way of explaining the moral synergy between doing and allowing harm, on which allowing a harm which we also do makes it harder to justify doing that harm, than it would be had we not also allowed it.

Recall *Assured Crushing*, discussed in §2.2. There, I suggested that Bob might need *some* justification to permissibly push the button. The story discussed in §4.1 could not explain why that would be so. If doing harm were only wrong insofar as it amplifies the wrongness of allowing harm, then it would not be wrong to any degree for Bob to harm Tara by pushing the button in *Assured Crushing*, since Bob does not allow any harm by doing so. So perhaps the wrongness associated with doing harm while allowing it may not be fully

explained by the fact that, in general, doing harm modulates the stringency of our duty not to allow harm. Perhaps, then, there is a distinctive wrongness associated with doing harm irrespective of also allowing it.

So, does Bob need some justification to push the button in *Assured Crushing*? If we have any moral rights, it is plausible to think that these include rights not to be harmed by others. Thus, it would seem, Bob does need *some* justification to push the button because doing so would infringe a right of Tara's not to be harmed.

Arguably, though, such rights-infringements can have varying degrees of moral weight. For example, using a pencil to painfully poke another's hand (without her consent) would infringe a right not to be harmed and so, I take it, is *prima facie* wrong. Using a sharp knife to stab someone's hand (without her consent) would also infringe this right, but it is *more prima facie* wrong. Why? Perhaps the explanation lies in a general feature of rights-infringements captured by Judith Jarvis Thomson's Aggravation Principle,[19] which we can put roughly like this: the worse it is for a victim that a right of hers is infringed, the more gravely the victim is wronged by that right infringement. Or rather, I want to consider a narrower formulation: the more *harmful* it is for a victim that a right of hers is infringed, the more gravely the victim is wronged by that right infringement.[20] When some harm that someone suffers is not one that we

[19] Thomson (1990: 267).

[20] We might also think that rights-infringements can also be rendered more grave by closing off greater benefits to their victims, though perhaps to a lesser degree than by being harmful to their

allowed to happen, the one who suffers such harm would have suffered the same total magnitude of harm no matter what we chose. Thus, when we do harm without allowing it, our doing the harm does not result in her suffering more harm than if we had acted differently. Cases like *Convergence*, I think, suggest something interesting about the Aggravation Principle, namely that the principle should be spelled out, not in terms of how much harm a victim would have suffered had the aggressor done nothing, nor in terms of how much harm would have been suffered had the aggressor not infringed the right, but rather in terms of how much harm arising from the rights-infringement that the victim *could* have been spared had the aggressor done something else that she was able to do.[21]

This explanation, then, combines two ideas: (i) that harm-doing infringes a right (that allowing harm would not in itself infringe), and (ii) that in general, rights-infringements carry more moral weight the more harmful it is for the victim that the right was infringed. When we infringe someone's right without allowing them to suffer harm, the victim could not have been spared any magnitude of harm no matter what we did. When we infringe a right while allowing harm, however, we act in such a way that the victim could have been spared some harm had we not infringed the right, meaning that it is harmful for

---

victims. I leave this open. I thank an anonymous referee for pressing me to be more explicitly sensitive to the distinction between harming and failing to benefit in discussing the Aggravation Principle.

[21] Cf. Norcross (2005: 165-166), in particular, his Bobby Knight case.

the victim that we chose to infringe the right rather than doing something else. Thus, if we grant both the Aggravation Principle and the assumption that doing harm infringes a right, we can readily explain how allowing harm to someone amplifies the moral weight of doing harm to her that contributes to this harm.

In *Assured Crushing*, Tara has a right against Bob that he not push the button (given that pushing the button would do harm to Tara). Thus, pushing the button is *prima facie* wrong to some (small) extent and hence needs some justification. However, if, unlike in *Assured Crushing*, pushing the button would make Tara suffer more harm in comparison to what could have happened had Bob done something else (e.g., pulling Tara out from under the boulder), the Aggravation Principle would kick in, increasing the gravity of the infringement of Tara's right that Bob not push the button were Bob to push it, and so increasing the threshold needed to justify pushing the button and doing harm to her.

Now, how does the story account for the difference between *Convergence* and *Divergence*? Although Bob does infringe a right not to be harmed by pushing the boulder in both *Divergence* and *Convergence*, this right-infringement takes more to justify in the latter case than the former because it leads to more harm for the victim (Victor) in *Convergence*, compared with how much he could have suffered had Bob acted in some other way that he was able to act, but does not lead to more harm for the victim (William) in *Divergence* in comparison with how much he could have suffered had Bob acted in some other way that he was able to act. Given that in *Divergence*, Bob does not infringe any right of *Victor's* no matter what he does, there is no right-infringement in *Divergence* that is nearly as grave as Bob's infringement of Victor's right when he

pushes the boulder onto him in *Convergence*. So, in *Convergence*, Bob commits a grave infringement of a right by pushing the boulder onto Victor, whereas in *Divergence*, he commits a minor infringement of the same kind of right by pushing the boulder onto William and, at the same time, also fails to save one person's life. Saving two people's lives is, I take it, a weighty enough consideration to justify such a minor infringement plus a failure to save one other person's life,[22] but it is not a weighty enough consideration to justify the grave right-infringement involved in lethally crushing a non-doomed person (as Bob does to Victor in *Convergence*).

5. Conclusion

We can relate to harms in a number of ways. Doing and allowing are at least two such ways, and I have argued that relating in one way to harm neither requires nor precludes relating to it in the other way. Far from being an unusual type of case, *most* cases of harm-doing are also cases in which the harm-doer allows the harm they inflict, thereby allowing some total magnitude of harm to be suffered.

[22] Compare with a case in which Bob must choose between (i) doing nothing, (ii) saving one person's life, and (iii) saving two people's lives in a way that involves pushing an unconsenting bystander out of his way. Although pushing the bystander in option (iii) infringes a right of the bystander's, this infringement carries a relatively small degree of moral weight, because it does not harm her significantly. I take it that Bob would be permitted to take option (iii) in such a case even though it involves an infringement of a right, because just like in *Divergence*, the right-infringement is not grave enough to outweigh the moral pressure exerted by the prospect of saving two lives.

Recognizing this fact serves as a steppingstone that allows us to see that there is a moral synergy between doing and allowing harm: either the moral weight of allowing some total amount of harm is amplified when one also does the harms that contribute to the magnitude of harm allowed to be suffered, the moral weight of doing harm is amplified when one also allows the same subject to suffer harm at the same time, or both.

I then offered two preliminary explanations of this moral synergy. On the first, doing harm amplifies the moral weight of allowing that same harm because it is, in general, harder to justify allowing harm when we have a special duty not to allow it, and doing a harm is a way of acquiring such a special duty. On the second, doing harm to someone is directly *prima facie* wrong because it infringes a right of hers. Given that rights-infringements are in general graver the more harmful they are for their victims, an infringement of this kind will be less grave when the victim would have suffered just as much harm regardless of what the aggressor did. Thus, harm-doing will carry less moral weight if the harm-doer doesn't thereby allow harm. On this story, allowing harm amplifies the moral weight of doing the harm in question.

These two stories are entirely compatible with each other. I leave it open whether we should accept one or both of them. There is also much more that could be said about the stories themselves. I leave it to future work to work out the relevant details.

For now, I conclude by suggesting how we might reframe the Doctrine of Doing and Allowing in light of the observed moral synergy and the two explanatory stories I proposed. We should refine the standard formulation of the

DDA by saying that it is only doing harm *while also allowing it* that is harder to justify than merely allowing harm. We should recognize that it is the convergence of two ways of relating to harm – *both doing* and *allowing* – that makes ordinary harm-doing so hard to justify.[23]

[23] I am greatly indebted to Peter A. Graham, whose extensive feedback and numerous discussions were instrumental to my writing this paper. I would also like to thank many others who provided extremely valuable feedback at various stages of developing this paper: Alejandro Pérez Carballo, Christopher Meacham, Maya Eddon, Sam Schechter, Firdaus Gupte, Cruz Davis, Jean Pierre Cordero Rojas, Justin Mooney, Justis Koon, and anonymous referees. Thank you all.

References:

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

Foot, Philippa (1967). The problem of abortion and the doctrine of double effect. *Oxford Review*, 5:5-15.

---- (1984). Killing and letting die. In Jay Garfield and Patricia Hennessey (eds.) *Abortion: Moral and Legal Perspectives*. Amherst: University of Amherst Press.

---- (1985). Morality, action and outcome, in Ted Honderich (ed.), *Morality and Objectivity*, London: Routledge and Kegan Paul, pp. 23-38.

Grice, Herbert Paul (1975). Logic and conversation, in Donald Davidson & Gilbert Harman (eds.), *The Logic of Grammar*. Encino, CA: pp. 64-75.

Hanna, Jason (2014). Doing, allowing, and the moral relevance of the past. *Journal of Moral Philosophy*, 12: 677-698.

Kagan, Shelly (1988). The additive fallacy. *Ethics*, 99(1): 5-31.

Liu, Xiaofei (2012). A robust defense of the doctrine of doing and allowing. *Utilitas*, 24 (1): 63-81.

Moore, George Edward (1903). *Principia Ethica*. Dover Publications.

Norcross, Alastair (2005). Harming in context. *Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition*, 123(1/2): 149-173.

Quinn, Warren (1989). Actions, intentions, and consequences: The doctrine of doing and allowing. *Philosophical Review* 98 (3): 287-312.

Rachels, James (1975). Active and passive euthanasia. In S. Cahn (ed.), *Exploring Philosophy: An Introductory Anthology*. Oxford University Press.

Thomson, Judith Jarvis (1976). Killing, letting die, and the trolley problem. *The Monist*, 59(2): 204-217.

---- (1990). *The Realm of Rights*. Harvard University Press: Cambridge, Massachusetts.

Unruh, Charlotte Franziska (2021). Letting climate change. *Journal of the American Philosophical Association*, 7(3): 368-386.

Woollard, Fiona (2015). *Doing and Allowing Harm*. Oxford University Press.