

## **Letting Climate Change**

*Forthcoming in the Journal of the American Philosophical Association; please cite published version*

Charlotte Franziska Unruh

University of Southampton

Cfu1e17@soton.ac.uk

Recent work by Ingmar Persson and Jason Hanna has posed an interesting new challenge for deontologists: How can they account for so-called cases of letting oneself do harm? In this paper, I argue that cases of letting oneself do harm are structurally similar to real-world cases such as climate change, and that deontologists need an account of the moral status of these cases to provide moral guidance in real-world cases. I then explore different ways in which deontologists can solve this challenge and argue that the most promising way to conceive of cases of letting oneself do harm is as non-standard cases of allowing harm, supplemented with an additional argument for the moral relevance of one's own agency. The upshot is that cases of letting oneself do harm are both more theoretically challenging and practically important than has been acknowledged.

Keywords: Doing/allowing, Harm, Deontology, Climate Change, Agency

Acknowledgements: I am grateful to participants at the International Society for Utilitarian Studies Conference 2018, Aart van Gils, Brad Hooker, and Fiona Woollard for excellent comments and discussion. I would like to thank anonymous reviewers for very helpful comments. This work was supported by the Arts and Humanities Research Council [AH/L503939/1].

## 1. Introduction

Most deontologists accept the Doctrine of Doing and Allowing (DDA), which says that doing harm is harder to justify than merely allowing harm, everything else being equal. The DDA can explain, for example, why killing one person is harder to justify than merely letting two others die.

Recent work by Ingmar Persson (2013) and Jason Hanna (2015a; 2015b) has posed an interesting new challenge for deontologists who accept the DDA (a qualification I will omit in the following). The challenge is to specify the moral status of so-called cases of letting oneself do harm.

(Poisoner) Earlier this morning, Agent deposited a dose of lethal poison into a teapot from which Victim drinks tea at the same time each afternoon. Unless Agent warns Victim now, Victim will drink the tea and die (Woollard and Howard-Snyder 2016; the original version of this case is found in Hanna 2015a: 678).

Failing to warn Victim would not be killing Victim. Rather, it would be failing to prevent Agent's previous behavior (putting poison in the teapot) from constituting a killing. By failing to warn Victim, Agent would *let herself do harm* (Hanna 2015a: 679). What is the moral status of cases of letting oneself do harm?

Consider a variant on (Poisoner). Assume that Agent can either warn Victim or save the lives of two innocent strangers, who are drowning in a shallow pond. Agent does not have the time to do both. Intuitively, Agent ought to save Victim's life (Hanna 2015a: 678). It seems to matter morally that Agent herself has created the threat that Victim faces. Let us call this intuition the 'self-other divide' (Persson 2013: 103).

Note that deontologists could explain the self-other divide by arguing that letting oneself do harm has the same moral status as doing harm. Just as it is impermissible to kill

one person to save two others, it is then impermissible to let oneself kill Victim to save two strangers.

Now consider a second variant on (Poisoner). Assume that *two* victims will drink tea out of the poisoned tea pot, unless Agent warns them in time. However, Agent can only reach the victims in time if she runs over (and thereby kills) an innocent stranger trapped on the road. Intuitively, Agent ought not to kill the stranger (Hanna 2015a: 681). It seems to matter morally whether the harming behavior occurs now or has been performed in the past. Let us call this intuition the ‘present-self focus’ (Persson 2013: 103).

Note that deontologists could explain the present-self focus by arguing that letting oneself do harm has the same moral status as merely allowing harm. Just as it is impermissible to kill one person to save two others, it is then impermissible to kill one person to avoid letting oneself do harm to two victims.

These explanations seem to conflict. This illustrates that the intuitions behind the self-other divide and the present self-focus point to two different ways of thinking about the moral status of cases of letting oneself do harm. These cases seem to occupy a curious status in between doing harm and merely allowing harm. The challenge for deontologists that arises from these cases is to provide a coherent account of letting oneself do harm that reconciles these intuitions.

If deontologists fail to explain intuitive verdicts about cases such as (Poisoner), this is bad, insofar as it is usually claimed to be an advantage of deontological moral theories that they can explain common-sense moral intuitions. This contrasts with views such as utilitarianism, which counter-intuitively imply that Agent ought to save the two drowning strangers instead of warning Victim, and that Agent ought to drive over the stranger on the road to save the two Victims.

My aim in this paper is twofold. The first aim is to highlight the ubiquity and practical importance of cases of letting oneself do harm. I do this by showing, first, that cases of letting oneself do harm are only a subset of cases involving complicated causal chains, all of which are intuitively different from standard cases of doing and allowing harm (section 2). Furthermore, such cases are not just a theoretical curiosity. Rather, they are widespread in real life – indeed, they are central to intergenerational ethics (section 3). Insofar as cases of letting oneself do harm challenge deontologists, they challenge the ability of deontologists to provide moral guidance in real-world cases such as climate change.

The second aim is to provide an analysis of cases of letting oneself do harm. I argue that recent defenses of the DDA which rest on a causal analysis of the distinction between doing and allowing harm, such as Fiona Woollard's, are incomplete if they do not provide an account of cases of letting oneself do harm. Deontologists can respond to these cases by understanding cases of letting oneself do harm as non-standard cases of doing harm, with an additional argument for why temporal nearness of the harming behavior matters. Alternatively, they can understand cases of letting oneself do harm as non-standard cases of allowing harm, with an additional argument for why our own agency matters. I argue that the second approach is more promising (section 4).

## 2. Letting Oneself Do Harm and Even More Complicated Cases

In this section, I argue that cases of letting oneself do harm are only one type of a whole family of cases with a complicated causal structure, whose moral status seems similarly unclear. The challenge presented by cases of letting oneself do harm is therefore much broader in scope than has been acknowledged.

Let us start by looking at the causal structure of cases of letting oneself do harm more closely. It will help to contrast them with standard cases of doing and allowing harm. A

natural way to understand the distinction between doing and allowing harm is in terms of the causal relation between an agent's behavior and a harmful outcome.

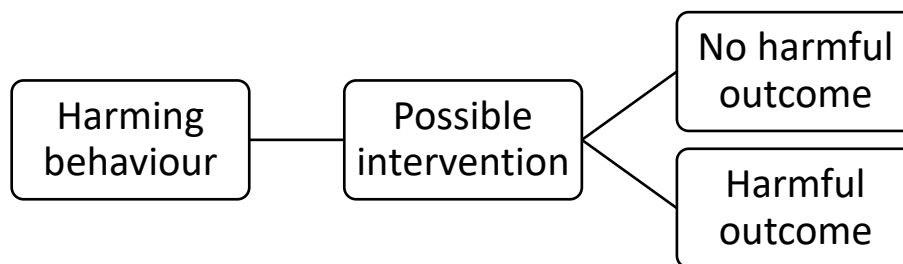
(Push) Agent pushes a boulder which rolls over Victim, crushing her to death.

(Non-Diversion) A boulder is rolling towards Victim. Agent could stop the boulder but does not do so. The boulder rolls over Victim, crushing her to death. (These cases are variants of cases discussed in Woollard (2015: 3) and Bennett (1993: 76–77).)

(Push) is a standard case of doing harm. Agent pushes the boulder (harming behavior) and Victim dies (harmful outcome). The harmful outcome follows directly from the harming behavior. (Non-Diversion) is a standard case of merely allowing harm. Agent stands still when the boulder rolls towards Victim (allowing behavior) and Victim dies (harmful outcome). If Agent had behaved differently (i.e. diverted the boulder), the harmful outcome would not have occurred. The DDA says that pushing the boulder in (Push) is harder to justify than failing to stop the boulder in (Non-Diversion), everything else being equal (Woollard 2015: 4; see also Foot 1967; Kamm 2007; Quinn 1989).

However, in the real world, causal chains leading to a harmful outcome are much more complicated than in (Push) or (Non-Diversion). In reality, causal chains are extended in time. They might involve additional or intervening actions or events. Whether the harmful outcome occurs might, in part, depend on these additional or intervening factors.

It will be useful for our discussion to think about cases of letting oneself do harm, such as (Poisoner), in a more abstract way as one type of such complicated cases. We can think of them as cases in which an agent is relevant to a harmful outcome through more than one behavior. I illustrate this in figure 1.



*Figure 1*

At the time of the harming behavior (for example, when putting poison in the teapot), an agent creates a threat now that she has the opportunity to avert later, or that she could have averted earlier. At the time of the possible intervention, an agent can allow harm that results from a threat that she herself has created earlier or will create later. For example, Agent can fail to warn Victim, with the result that Agent's previous action harms Victim. As explained above, such a behavior seems harder to justify than merely allowing harm (this intuition is captured by the self-other divide), but easier to justify than doing harm (this intuition is captured by the present self-focus).

Hanna brings up a further complication. Consider

(Two Buttons) Last week, Agent initiated a threat that is now about to kill five people. Five minutes ago, Agent realized that she could save the five by pressing button 1, but that pressing button 1 would also kill one. She has pressed button 1. Now, she has the opportunity to press button 2. Pressing button 2 will interfere with the reactions triggered by pressing button 1 and will prevent both the saving of the five and the killing of the one. (This is a shortened version of a case given by Hanna (2015a: 689–690).)

The present self-focus implies that Agent's pressing button 1, thereby initiating the reaction, is morally wrong. But what should deontologists think about pressing button 2?

According to Hanna, they 'would presumably claim that Agent is morally required to press Button 2 and terminate the reaction. This is because it is very counterintuitive to hold

both (a) that Agent is morally forbidden from initiating the reaction and (b) that, if he [Agent] does initiate the reaction, he is permitted to let it continue when he could easily stop it' (2015a: 690).

Perhaps Hanna is right. The assumption that we should revert our wrong actions, if we can, seems intuitively plausible, and indeed implies that Agent should press button 2.

However, *pace* Hanna, one can argue that pressing button 1 has changed the decision situation that Agent faces when considering whether to press button 2. After pressing button 1, the five are no longer under a threat, and it is impermissible to recreate this threat. On this view, Agent should not press button 2.

Be that as it may, it is not clear how the DDA would lend support to either view. Agent has initiated a threat to the five, and a threat to the one. She would let herself do harm in pressing button 2, and she would let herself do harm in not pressing button 2. Insofar as the moral status of letting oneself do harm is unclear, the moral status of cases such as (Two Buttons) is similarly unclear.

However, deontologists might object to this claim. They might argue that it is not the business of the DDA to give verdicts in cases like (Two Buttons), and perhaps (Poisoner). Rather, the moral status of these cases is determined by considerations outside the scope of the doing/allowing distinction.

I briefly defend the claim that at least some accounts of the DDA would be incomplete without an account of cases of letting oneself do harm in section 4. In the meantime, let me offer some reasons to think that at least what seem like the most obvious candidate considerations cannot account for the moral status of cases such as (Two Buttons).

Can deontologists just appeal to the numbers? After all, surely letting oneself do harm to five is worse than letting oneself do harm to only one, everything else being equal. So, Agent should not press button 2. However (and ignoring any principled reservations that

deontologists might have about deciding by the numbers in this case), appeal to the numbers does not get us all the way. Here is why. If Hanna is right and we have stringent requirements to revert our wrong actions, then deontologists should press button 2, even if this saves the smaller number. If, however, prohibitions against recreating threats are stronger, then deontologists should presumably not press button 2, even if this would save a greater number.

To be clear, I do not suggest that the numbers do not matter. I make the different claim that two plausible lines of deontological reasoning about cases such as (Two Buttons) seem to recommend either pressing or not pressing button 2, independently of the numbers that are at stake. Importantly, both views might recommend saving the smaller number. This seems to indicate that deontologists should not appeal to numbers in cases like (Two Buttons).

Can deontologists appeal to the distinction between intending and merely foreseeing harm? I think that appealing to the role of intention is unlikely to solve the problems raised by the cases discussed in this section (I am here in agreement with Hanna (2015a: 693)). The problem seems to arise in just the same way when we assume that Agent has performed all relevant behaviors (pushing all the different buttons) intentionally as when Agent has intended none of the deaths. Moreover, whether Agent has acted intentionally or negligently does not seem to change intuitive verdicts about cases such as (Poisoner). Whether Agent has put the poison in the teapot intending to kill Victim, being merely negligent, or finding the only safe place to hide it from the Mad Poisoner does not seem to change the intuition that Agent ought to save Victim rather than five strangers, and that Agent should not drive over one person to save two potential victims.

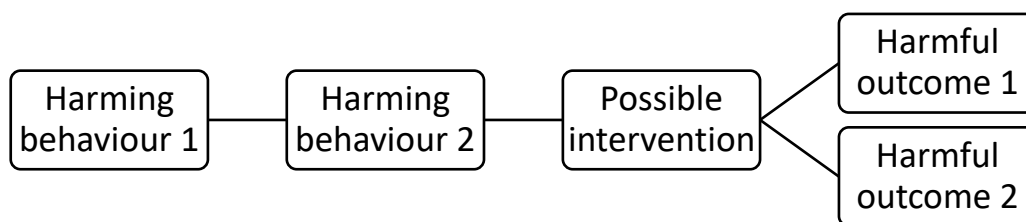
Neither appeal to numbers nor appeal to the relevance of intention can provide a clear account of the moral status of complicated cases. I therefore maintain that insofar the DDA does not provide a clear account of the moral status of cases of letting oneself do harm, the



same can be said about cases with a more complicated causal structure, such as (Two Buttons).

I will now argue that (Two Buttons) is just one example of a family of cases in which agents can intervene in the causal chain leading to harm more than once. There are many possible cases that share a similar structure. An agent might have created, or know that she will create, multiple threats. She might now be able to avert some, but not all, of them. Or she might have previously allowed harm that she now has a second chance to avert, but only at the cost of creating a further threat.

To see this, it is helpful to consider the causal structure of such cases. In these cases, an agent is relevant to more than one possible outcome through a single intervention behavior, as illustrated in figure 2. For example, in (Two Buttons), whether Agent presses button 2 decides which of her earlier behaviors will lead to harm.

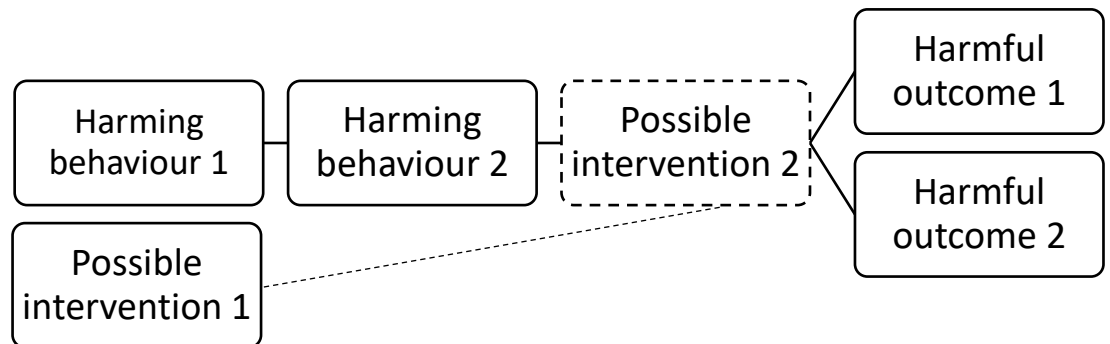


*Figure 2*

To make matters worse, cases like (Poisoner) and (Two Buttons) do not exhaust the possibilities of complicating causal chains. Consider

(Third Button) Agent anticipates well ahead of time that she will be in the situation described in (Two Buttons). She is uncertain about what she is supposed to do in that situation. She can now press button 3. Pressing button 3 disables button 2, so that once Agent has pressed button 1, no further intervention will be possible (and she will be spared the resulting moral quandaries!).

(Third Button) is an example for a case in which an agent acts in a way that influences whether, and if so how, she will be able to intervene in the causal chain leading to a harmful outcome more than once, as illustrated in figure 3.



*Figure 3*

What should deontologists think about cases such as (Third Button)? If we should revert a wrong decision when we can do so, then it seems equally plausible that we should not prevent ourselves from being able to revert a wrong decision in the future. Therefore, if one believes that Agent ought to press button 2, it seems that one should also believe that Agent ought not to press button 3.

However, if one believes that Agent ought not to press button 2, it seems that one should also believe that Agent is at least permitted to press button 3, since doing so would only make a morally wrong future option (pressing button 2) unavailable.

Again, the DDA does not seem to support either verdict. I have said above that Agent lets herself do harm both by pressing button 2 and by not pressing button 2. Therefore, when pressing button 3, Agent prevents herself from letting herself do harm later. When refraining from pressing button 3, Agent lets herself let do harm later. Since the moral status of pressing button 2 was unclear to begin with, the moral status of pressing button 3 seems even less clear.

There are several possible cases with a similar structure as (Third Button). For example, an agent can now initiate a threat while making sure that she will have an opportunity to avert the harmful outcome later. Or she can do something now to prevent a harmful outcome, while making sure that she can later undo the prevention, in which case the harmful outcome would nonetheless occur. (Third Button) is therefore just one of a family of complicated cases which share causal features with standard cases of doing and allowing harm. The moral status of these cases is unclear.

We can conclude that the range of cases in which the DDA does not deliver a clear verdict goes far beyond cases like (Poisoner). The possibilities for further complication are endless.

### 3. Complicated Cases in the Real World

These cases are admittedly confusing, and my anecdotal evidence suggests that few people have clear intuitions about them. All examples discussed above, from (Poisoner) to (Third Button), are very complicated and highly artificial. Given this, one might think, perhaps it is not so bad after all that we cannot come up with a theoretical account of such cases.

Many people can live very comfortably in the presence of unsolved quibbles in the periphery of their preferred moral theories. And even those who would prefer their moral theory to apply even to the most fantastic hypothetical cases might think that cases of letting oneself do harm do not make the priority list of problems to which defenders of the DDA should devote time and energy.

Moreover, in real life, we do not often seem to find ourselves in scenarios such as (Poisoner). Everyday harm doing seems to be much more straightforward: We do something, and the harm is done, usually before we had another opportunity to intervene.

However, there is a set of real-life cases that is structurally similar to the complicated cases discussed above. These real-life cases frequently arise in intergenerational ethics, specifically in cases that involve doing harm to future generations. (I assume that we can harm future generations in morally relevant ways a claim that is plausible but controversial (see Campos 2018: 3–4).)

In future generation cases, by definition, we deal with the long-term consequences of our present behavior. Present actions seldom have direct and inevitable effects that will obtain in a few centuries. Usually, there is something that we can do, now or later, to intervene in the causal chain leading to harmful outcomes in the (far) future. If we refrain from doing so, we let ourselves do harm. For an example, think about climate change, and ideas that scientists have come up with to prevent its harmful impacts: from reducing carbon emissions in the first place, over developing technology to lower the level of carbon already in the atmosphere, to mitigating the adverse effects caused by global warming.

Of course, it is not straightforward who the ‘we’ in such cases refers to. Possible candidates include: ‘We’, the developed nations, which continuously emit carbon dioxide, causing climate change. ‘We’, nation so-and-so, which has in the past emitted more carbon dioxide than our commitments permit. ‘We’, city so-and-so, that will not put into force regulations to ban plastic straws in restaurants. ‘We’, the Joneses, who still eat meat and drive cars. There are countless complexities arising here, regarding collective and shared agency, the notion of causation, and collective and institutional agency over time.

However, consider the case of a government that provides extensive funding for coal-fired power stations. It seems intuitively plausible that the government is doing harm in contributing to climate change. It also seems plausible that, if the government 50 years from now is presented with an opportunity to combat climate change, they would let themselves do harm if they did not avert climate change. Even if the persons constituting the government

will have changed, in a sense the government can be ascribed institutional agency over time. It is in an important sense the same government that has funded the power stations, as has failed to combat global warming.

More generally, the following assumptions seem plausible: At least some collective and institutional agents such as governments, supranational organizations, companies, the inhabitants of village X, and perhaps even groups such as the ‘Millennials’, or the global affluent, possess agency over time in a relevant sense. They can therefore let themselves do harm to future generations. Because of the large time gap between harming behavior and harmful outcome, these agents will often have numerous opportunities to intervene with their own actions.

To illustrate, consider a climate change example that is structurally similar to (Poisoner):

(Climate Change I) In the past, we have emitted carbon dioxide in the atmosphere. If we do not take action now to mitigate climate change, future generations will suffer from, amongst other effects, a rise in infectious diseases.

Similarly, the following case is structurally similar to (Two Buttons):

(Climate Change II) Climate Change will lead to a future rise in infectious diseases. We initiate mitigation through increased use of green technology, even though this will inevitably cause comparable harm – perhaps a future rise in infectious diseases due to industrial activities. However, shortly before these industrial activities start, we begin to doubt the morality of our action and consider stopping green technology.

The following case is structurally similar to (Third Button):

(Climate Change III) Climate Change will lead to a future rise in infectious diseases. We can now decide to adopt regulations that prevent us from stopping green technology, once we have decided to go for it. This would mean that after our

decision to mitigate by increased use of green technology, we would not be able to revert this decision and industrial activities would inevitably cause a huge rise in infectious diseases.

It is beyond the scope of this paper to consider the many fascinating issues that arise from these cases. One particularly interesting question is who qualifies as an agent that can let themselves do harm (and bear responsibility for this behavior). While I cannot hope to conclusively answer this question here, I offer some considerations which underline my claim that cases of letting oneself do harm are more common than one might think.

Agents who can let themselves do harm need to be (i) able to ‘do’ or ‘allow’ harm in the first place, and (ii) persist over time in a way that makes it possible to ascribe them responsibility for earlier actions. Regarding the first criterion, it seems plausible that in order to be able to do or allow harm, an agent must possess some causal power in the sense that their behavior potentially makes a difference to the outcome. For example, one individual presumably cannot, on their own, make a perceptible difference to the overall level of carbon dioxide emissions. However, the signatories of the Kyoto Protocol can.

Regarding the second criterion, to persist over time in the relevant sense, an agent must be clearly identifiable as the agent performing both the allowing and the doing behavior. For example, a city council might count as letting themselves do harm when they refrain from replacing the faulty train switch boards that they bought cheaply years ago, leading to runaway trolleys on the rails (and pedestrians being pressured into putting their moral views in action!). The interest group that has the opportunity to put pressure on the city council to replace the switch boards, but fails to do so, might count as allowing harm, but not as letting themselves do harm (assuming that they had not been involved in the previous decision to buy the cheap switch boards). These considerations should lead us to resist the temptation to argue that, since past generations have caused climate change, present generations fulfil their

obligations not to harm by keeping present emissions low, thus not adding to the burden. Insofar as the agent who caused climate change in the past is best described as an agent existing over time (e.g. ‘developed nations’), they still count as letting themselves do harm if they refrain from mitigating, even if they did not produce additional emissions.

One of the central aims of moral theorizing is to arrive at action-guiding conclusions. An essential part of that is to figure out how moral concepts and theories apply to real-world cases. If that is correct, then finding out how to apply the DDA to real-world cases in intergenerational ethics – including complicated cases such as those described above – is an urgent task. Indeed, I think the urgency of that task is underappreciated in contemporary moral philosophy.

To illustrate this, let me hint at the implications that might arise in the absence of an account of the moral status of complicated cases. If complicated cases are much more common in future generation cases than in present generation cases, and if we are therefore uncertain about the moral status of many future generation cases, then it might seem natural to conclude that moral obligations towards future generations often differ from moral obligations towards present people. This is because, as long as the moral status of complicated cases remains unclear, we cannot compare them against standard cases in ways that would be necessary to determine our moral obligations. For example, when facing a decision between an option that involves a risk of standard harmdoing, and an option that involves a complicated case, it would be unclear whether, why, and how much one of these options would be harder to justify than the other. Without an account of complicated cases, this might give us a reason to reject either the arguably widespread belief that our obligations towards future generations are relevantly like obligations to those presently alive, or the distinction between doing and allowing harm itself – at least regarding its usability for actual decision making.

However, we need not accept these implications just yet. What I have said so far is, I believe, less of a challenge than a pointer towards an area that is currently underexplored. This is what deontologists who accept the DDA might say: ‘Clearly, what you call the ‘complicated cases’ are not clear-cut cases of either doing or allowing harm. These cases are what one might call non-standard cases. They are subsets of harm doings, or perhaps harm allowings, and they are different from those standard cases in morally relevant ways. If we can specify these differences in terms of causal relations to harm, this is great – spelling this out in more detail might tell us more about the dynamics that make a certain causal relation to harm easier, or harder, to justify.’

If this is right, then my argument in the last section does not point to a flaw in deontological moral theory. Rather, it draws attention to an area where the otherwise abundant literature on the DDA remains surprisingly silent and points out the urgency to explore this area further. In the remainder of the paper, I will explore some possible directions that the defender of the DDA could take here.

#### 4. The Moral Status of Cases of Letting Oneself Do Harm

Let me first, however, motivate the claim that cases of letting oneself do harm are a challenge in particular for defenders of the DDA. I will do this by providing a reason to think that the DDA would be incomplete without a causal analysis of complicated cases, and an account of their moral status.

Is there a moral difference between doing and allowing harm? Recent defenses of the DDA provide a two-step strategy for answering this question in the affirmative. First, they analyze the distinction between doing and allowing harm. Second, they argue that the underlying nature of this distinction makes it morally relevant (Woollard and Howard-Snyder 2016). Most recently, Fiona Woollard (2013; 2015), following Philippa Foot (1967),



distinguishes doing from allowing harm in terms of the causal relation between an agent's behavior and a harm. Woollard's first-step analysis of the doing/allowing distinction relies on facts about how the agent's behavior is relevant to the harm in question, where facts about the agent's behavior form part of a causal sequence leading to harm.

According to Woollard (2015), an agent's behavior is *doing* harm if and only if a fact about the behavior is part of a sequence leading to harm. An agent's behavior is *merely allowing* harm if and only if a fact about the behavior is relevant to, but not part of, a harmful sequence. Only *substantial* facts – roughly, facts that are either positive, or against normal presuppositions – can be part of a sequence. It is worth noting that the fact that the agent remains still can sometimes be a positive fact about behavior. For example, consider a case in which an agent is knocked over on top of a child. If the agent remains perfectly still, the child will asphyxiate. The fact that the agent remains perfectly still tells us exactly how the agent behaved. This fact can therefore be part of a sequence leading to harm (Woollard 2015: 49–50; see also Bennett 1995: 99).

Having analyzed the distinction between doing and allowing harm, Woollard then goes on to argue that the distinction is morally relevant. Roughly, in her view, the DDA protects moral agents from undue imposition. It provides this protection by requiring agents to refrain from doing harm and by permitting them to allow harm to others.

However, these are exactly the kind of facts that seem to distinguish letting oneself do harm from merely allowing harm. If defenders of the DDA think these facts make a moral difference, then they should be able to say whether, and if so, how they do so in the case of letting oneself do harm (and, indeed, more complicated cases).

Letting oneself do harm as in (Poisoner), prioritizing saving some of one's own victims over others as in (Two Buttons) and barring oneself from possible future interventions as in (Third Button), are behaviors that are clearly relevant to the harm that

occurs. However, it is less clear *how* the agent's behavior is relevant to the harm in these cases: Is it in a doing way, in an allowing way, or in a yet-to-be-specified third way? In the remainder of this section, I will explore these possibilities in turn. I will focus on relatively simple cases of letting oneself do harm, such as (Poisoner). An account of the moral status of these cases would constitute a first step towards a more comprehensive account and is likely to shed light on more complicated cases.

#### **4.1 Letting Oneself Do Harm as Doing**

If Agent does not warn Victim in (Poisoner), she will have done harm. Victim will die as a result of Agent poisoning the tea, rather than being saved. Agent will have killed Victim! So perhaps one might think that letting oneself do harm is best seen as an instance of doing harm.

But how can a failure to warn Victim constitute a doing behavior? One *prima facie* plausible way to think about this is to say that one should understand Agent's harm doing as extended in time. The extended harm doing contains all behaviors relevant to the harm: putting the poison in the teapot and failing to warn Victim.

As pointed out above, however, the view that letting oneself do harm has the moral status of doing harm fails to explain the present self-focus. Recall that this was the intuitive claim that doing harm now (e.g., driving over Stranger) is much harder to justify than letting oneself do harm now (e.g., warning Victim).

However, perhaps deontologists can provide a separate defense of the present self-focus. And indeed, it seems like there is a good justification for focusing on the agent's present behavior. This is, roughly, what the defender of the present self-focus might say (arguments along these lines have been given by Thomson (1985: 1415) and Woollard (2014: 276)): 'As human beings, we exercise our agency at present. We typically have immediate

access to current deliberations, and immediate control over current choices, given the available options. Our control over our present behavior is typically greater than our control over our past or future behavior. This is why our present choices carry special moral weight. Initiating a harmful sequence now is therefore harder to justify than not stopping a harmful sequence that you have initiated.’

This line of argument can explain why threats that we create now matter more (as in Poisoner). However, it cannot explain why one of two threats that we have created in the past seems to matter more than the other (as in Two Buttons).

It might be tempting for the defender of the present self-focus to argue that temporal nearness of the harming behavior matters – perhaps it does so indirectly, through the importance that we attach to the present decision-making.

However, even if such an argument could be made, it would not be able to account for all cases that are relevantly like (Two Buttons) (and, for that matter, (Climate Change II)). Hanna suggests that our intuitions about what matters morally track ‘temporal proximity, or [...] how close to the present the harming behavior would occur’ (2015a: 696).

However, *pace* Hanna, it does not seem to be true that our intuitions track temporal nearness. Rather, they seem to be tracking temporal order of the harming behaviors. Consider, again, (Two Buttons). Imagine that Agent knows she will initiate a threat to the five. She also knows she will then press button 1. She can now disable button 1, such that when she presses it later, nothing will happen.

If temporal nearness of the harming behavior mattered morally, we would expect that, other things being equal, we ought to interrupt the harmful sequence that we will initiate sooner, and that Agent should therefore not disable (and indeed, should later press) button 1.

However, intuitively, just as in the original (Two Buttons) case, it seems wrong to press button 1, and thus permissible to disable button 1 now. When Agent disables button 1,

she prevents herself from initiating a harmful sequence to the one. She does not prevent herself from initiating the (earlier) threat to the five.

If this is correct, then it is misleading to say that our intuitions track the temporal *nearness* or *proximity* of harmful behavior. Rather, they seem to track the temporal *order* of harmful behavior, or threat initiation: It is more objectionable to let oneself initiate a new threat than it is to refrain from interrupting a threat that one has previously caused. It is at least unclear whether and if so, how, the argument that the ‘present tense matters’, as Thomson (1985: 1401) puts it, can account for the temporal order of behaviors.

#### **4.2 Letting Oneself Do Harm as Allowing**

So maybe letting oneself do harm is better described as allowing harm. However, as pointed out above, the self-other divide seems to indicate that this is mistaken. Letting oneself do harm is, everything else being equal, harder to justify than merely allowing harm. In (Poisoner), the fact that it was Agent who put the poison in the teapot makes a difference.

One might think that this is not surprising. Letting oneself do harm, after all, is not *merely* allowing harm. Therefore, one might object, we should not expect its moral status to be equivalent to *merely* allowing harm. The objector could then point out that some instances of allowing harm are harder to justify than other instances of allowing harm.

There are at least three factors that can change the moral status of clear cases of allowing harm. First, facts about the relationship between Agent and Victim might matter. For example, not saving Victim might be harder to justify if Agent is Victim’s friend or has promised to save Victim. Second, facts about the situation might matter. For example, not saving Victim might be easier to justify if it would be very difficult for Agent to do so. Third, facts about Agent’s epistemic or motivational state might matter. For example, allowing harm

might be easier to justify if the harm is a foreseen, but not intended, consequence of Agent's behavior.

However, the intuition that letting oneself do harm is harder to justify than merely allowing harm persists even when the factors mentioned above are changed. It therefore seems that whether agents let themselves do (or merely allow) harm is not a proxy for a morally relevant difference, but itself possesses moral relevance.

Agent intuitively ought to save Victim rather than Stranger, even when facts about the situation, the relationship between Agent and Victim, and Agent's knowledge, intentions and motivations change. To illustrate, Agent should save Victim even if Victim is a stranger, even if Agent did not maliciously intend the poisoning, and even if Agent was pressured into poisoning Victim.

While it is entirely plausible that these factors matter morally, they do not explain the difference between cases of letting oneself do harm, and cases of merely allowing harm.

Fortunately, there is an alternative, intuitively plausible explanation of the difference between cases of letting oneself do harm and cases of merely allowing harm. The difference lies in the way in which Agent's behavior is relevant to harm. What matters is that, in addition to Agent's present allowing behavior, in cases of letting oneself do harm there are facts about Agent's past or future behavior that are relevant to the harm in a doing way.

In other words, Agent is relevant to a harm through two behaviors, one of which is clearly a doing behavior. If Agent does not warn her victim in (Poisoner), she thereby ensures that she will have done harm. She allows her past behavior of poisoning the tea to constitute a harm doing.

To be more precise, it seems that letting yourself do harm is different from merely allowing harm in three ways: (i) facts about the agent's behavior (distinct from the present behavior) are relevant to the harm in question; (ii) these facts potentially qualify the distinct

behavior as harm doing; and (iii) whether they do so depends on facts about the agent's present behavior.

However, taken by itself, the mere fact that one's present behavior can influence the effects of one's past or future behavior does not provide an explanation for the intuition behind the self-other divide. Nor does the mere claim that in cases of letting oneself do harm, we need to include all potentially relevant past and future behaviors in our evaluation. Surely, we would want to be able to evaluate an agent's behavior before the last opportunity for the agent to interfere in the causal chain leading to harm has passed and the harm has occurred. The DDA would not be very useful if it told us to refrain from the moral evaluation of harming behaviors, effectively until the harm has materialized.

A more promising way to explain why letting oneself do harm is harder to justify than merely allowing harm is to provide a more robust defense of the claim that our own agency matters. Such a defense might start by pointing out that as human beings, we are essentially embodied. This is, roughly, what the defender of the self-other divide might say: 'Our relationship to our own body is fundamentally different from our relationship to others' bodies. Our relationship to actions that we ascribe to ourselves is fundamentally different from our relationship to others' actions. We are responsible in a unique way for actions that we ascribe to ourselves. Without a sense of responsibility for our own agency, it would be hard to see how we could think of ourselves as individual agents, who deliberate, decide, act and interact with agents who are clearly separate from ourselves. This is why our own agency carries special moral weight. Not stopping a harmful sequence that you have initiated yourself is therefore harder to justify than not stopping a harmful sequence that you have not initiated.'

To be clear, I do not suggest that sequences of acts should never be evaluated as a whole. How to individuate different behaviors that form parts of sequences is a complicated

topic, which I do not discuss here. However, finding an account of how to individuate such behaviors is a yet-to-be-solved challenge for all theories that operate with causal sequences, and is therefore not a challenge specifically for my argument.

The thought that responsibility for our own agency is necessary for us to see ourselves as individual agents is inspired by Susan Wolf's (2001: 12–13) explanation for responsibility for harm arising from bad moral luck. We can back up this reasoning with considerations of autonomy or self-ownership. Woollard (2014: 275) suggests that it is more objectionable to let oneself do harm, rather than just allow it, because one is (part of) the threat in the former case but not in the latter. A duty to aid seems to demand less when the agent herself is part of the threat. The agent's body is not imposed upon in the same way and her self-ownership rights are not infringed upon in the same way as when she is not causally connected to the harm in question.

However, more needs to be said to fully explain and justify the self-other divide. Defenders of the self-other divide need to explain why our own agency should matter enough to override considerations about the overall consequences of actions. In other words, why should our own agency matter *that much*?

The question is therefore not only why responsibility for our own agency should matter morally, but also why it should count for as much as our intuitions seem to suggest. This becomes especially apparent when considering cases under uncertainty such as

(Potential Poisoner) As in (Poisoner), but this time, Agent is not sure whether the white stuff she put in the tea earlier was arsenic or sugar. She can now either save five strangers, who, as she knows, have ingested arsenic, or hurry back to pour away the tea.

Intuitively, Agent is at least permitted, if not required, to let the five strangers die to ensure that she will not have poisoned the five. This seems to be true whether Agent is 95%, 85%, or

even only 75% certain that she has put poison in the tea. The self-other divide can vindicate such intuitions.

However, it is hard to see how such special concern with one's own actions can be underpinned morally, especially since the motivation for the intuition behind the self-other divide can be called into question: One might say that our intuitions underpinning the self-other divide stem from an ultimately indefensible desire to keep our own hands clean. This concern is also raised by Persson, who thinks there might be a worry of 'repulsive moral self-indulgence [...] why be especially concerned about your own right-violations rather than the right-violations of all people, in proportion to the stringency of the rights violated?' (2013: 110).

The challenge for defenders of the self-other divide, then, is to provide a plausible justification that is sufficiently strong to explain our intuitions in the Poisoner cases. In the following, I will sketch one way of providing such a justification, based on the idea that we have special obligations towards those whom we have wronged in the past, or will wrong in the future, by imposing a risk of harm on them (thanks to an anonymous reviewer for pressing me on this point).

This justification of the self-other divide is based on a line of reasoning that is discussed, but ultimately rejected, by Hanna (2015a: 685–86). Hanna suggests that defenders of the self-other divide could appeal to considerations from restorative justice. It is plausible that we can wrong others by imposing threats of harm on them, even if the harm will only materialize at some point in the future. By putting the poison in the teapot, then, Agent wrongs the five. Such wronging, however, gives rise to special obligations. Hanna points out that it is generally thought that we have special obligations to 'mitigate, prevent, or offset the harms associated with one's wrongful behavior'. Obligations to aid our past victims might just be a particular instance of these broad special obligations (2015a: 686). This explains



why our own agency matters morally. We ought to be more concerned with our own wrongful behavior, because our own behavior gives rise to special obligations towards those whom we have wronged. Moreover, these special obligations seem to have the required strength: they require agents to compensate their own victims over aiding strangers, but do not permit killing innocents in order to prevent or mitigate harm from one's potential victims (2015a: 686).

Hanna argues that the argument from restorative justice cannot account for cases of future harmdoing. He gives the example of an agent who has a condition that, if untreated, will cause violent sleepwalking. The drugs for curing the condition are expensive, and the money could prevent much more harm when donated than the agent would prevent by curing herself from the condition. Intuitively, the agent nonetheless ought to buy the drugs, to prevent herself from unwitting violent behavior in the future. But considerations from restorative justice seem irrelevant, as the agent has not yet wronged anyone (Hanna 2015a: 687–688).

However, I suggest that once we accept that we have special obligations towards those whom we have wronged in the past, we should also accept that we have special obligations towards those whom we will foreseeably wrong in the future. To illustrate, imagine an agent who is prone to drunk driving. When going out for a drink, this agent ought to take precautions to prevent herself from doing harm in the future, for example, by pre-ordering a cab. She ought to take these precautions, even if doing so is expensive (for example, if the ride home costs more than saving a life by donating to an effective charity). A natural explanation is that if the agent does not take such precautions, she will foreseeably wrong those endangered by her future behavior. Similarly, the violent sleepwalker ought to buy the drug, to prevent herself from wronging others by imposing a risk of harm on them.

One might object that we cannot have direct obligations towards those whom we have not wronged yet. However, it strikes me as plausible that we can have such obligations in situations in which we can foresee that we will wrong others in the future. For example, there does not seem to be a morally relevant difference between a case in which an agent has just poisoned a drink, and a case in which the agent knows that she will do so in five minutes (perhaps because of some compulsion), if she does not pour the drink away now. It seems that the agent in both cases has special obligations towards the drink's owner to pour the drink away.

However, my argument does not depend on this claim. We might still have indirect obligations to prevent ourselves from wronging others, as we would then incur direct special obligations. For example, even if the violent sleepwalker does not have direct obligations towards her potential future victims, she might have indirect obligations to prevent a situation in which she incurs special obligations by wronging her potential future victims. (Failing to donate would not ground special obligations towards those she has failed to aid.)

Another objection is that wrongdoing cannot give rise to special obligations if the wrongdoing was blameless, because agents are not morally responsible in those cases. I can see two ways in which the defender of the self-other divide might respond here.

One response would be to concede that Agent has no, or weaker, special obligations towards those she has wronged blamelessly. However, this seems to contradict intuitions about such cases. Assume that Agent was in no way at fault when putting arsenic in the sugar jar (one of the victims labelled the jar incorrectly). It still seems that as soon as Agent learns this fact, she should drive back to save the tea pot victims, rather than saving the strangers.

These remarks suggest a second response: The self-other divide strengthens reasons against letting ourselves do harm, regardless of whether we are blameworthy for our doing. Bernard Williams has famously described that we often ascribe some responsibility to agents

who, through no fault of their own, commit a wrong (such as the lorry driver who, despite taking all precautions, runs over a child) (Williams 1981: 28). We might think that a similar story can be told about the blameless poisoner. On this view, like the lorry driver, the unwitting poisoner should take responsibility for her action, and thereby incur special obligations towards the victims. (Of course, her blamelessness will likely change the overall moral evaluation of her behavior.)

A final problem concerns where we should draw the line between our own agency and external factors. Persson (2013: 95–96) describes a case in which Agent can feel the onset of a spasm, which, if it runs its course, will cause Agent to pull the trigger of a gun. According to Persson, by failing to suppress the spasm, Agent lets herself do harm. He argues, further, that it seems hard to justify why we should be more responsible for allowing harm that results from our involuntary bodily movements than for allowing harm that results from external factors: ‘The fact that a twitch is internal rather than external to us cannot make any moral difference’ (Persson 2013: 105).

In response, deontologists might question whether involuntary movements like spasms are things that we ‘do’ in the relevant sense. Bennett (1995: 110) seems to take this line, while noting the peculiarity of the spasm example, which he describes as a case of allowing harm in which the agent’s behavior is positively relevant to the upshot. Alternatively, deontologists might agree with Persson that involuntary movements are things that we do but insist that this wide sense of doing is morally significant. For example, Woollard argues that agents have a special relationship with their own body, which grounds special responsibility for what our bodies do (2013: 331; see also 2015: 192).

It is beyond the scope of this paper to settle this matter. However, doing so is not necessary for my purposes. I am interested in the moral status of cases of letting oneself do harm. Whether involuntarily movements can constitute doing harm (and, therefore, whether

allowing such movements can constitute letting oneself do harm) is a further question which does not pose a *special* challenge to the discussion of the moral status of cases of letting oneself do harm.

To sum up, I suggested that agents should be especially concerned with their own wrongdoing in virtue of the special obligations that they have towards those whom they have wronged in the past or know they will wrong in the future. These special obligations give agents moral reasons to prevent harm resulting from their own past or future doings. Absent any confounding reasons, these moral reasons are weaker than the reason against doing harm (agents ought not to kill in order to prevent themselves from killing), but stronger than the reason to prevent harm that is unrelated to the agent (agents ought to save their own victims over others).

### **4.3 Letting Oneself Do Harm as ‘Dallowing’**

For the sake of completeness, let me briefly discuss a final option. On the face of it, letting oneself do harm does not seem to be an instance of either doing or allowing harm. If we only consider Agent’s present action, then letting oneself do harm is clearly an instance of *allowing* harm. However, if we consider an agent’s behavior over time, then letting oneself do harm is clearly (at least part of) a harm *doing*.

With that in mind, let us turn to a third option. One might be tempted to classify cases of letting oneself do harm as belonging to a third category. This is what one might say: ‘Letting oneself do harm is in relevant ways like merely allowing harm, even if they are not equivalent. Letting oneself do harm is in relevant ways like doing harm, even if they are not equivalent. Letting oneself do harm shares morally relevant features with both doing and allowing harm, and thus the moral status of letting oneself do harm lies in between those.’ These cases would then form a distinct moral category by themselves.

If this reasoning is correct, then cases of letting yourself do harm constitute a distinct set of cases whose moral status is somewhere in between doing and allowing harm. For lack of a better word, we might say that letting oneself do harm would then correctly be described as an instance of ‘dallowing’.

However, the defender of the DDA would need to say more about this account. What does it mean to say that a ‘dallowing’ behavior has an ‘in between’ moral status? Is dallowing harm much harder to justify than merely allowing harm, or just a little bit harder? Is it much easier to justify than doing harm, or just a little bit easier? And how can the answer to these questions be defended? It seems that this option opens more questions than it solves.

Moreover, I think that deontologists have principled reason to resist the temptation to introduce new categories of cases alongside doings and allowings. Limiting the scope of cases that can be classified as either doing and allowing harm also limits the explanatory power of the doing/allowing distinction. It seems preferable to be able to classify cases of letting oneself do harm as either doing harm or allowing harm and spelling out the ways in which they differ from paradigm cases of doing and allowing harm, rather than opening up new categories, and with it the need for new and expanded conceptual and moral frameworks.

## 5. Conclusion

With powerful technologies and scientific methods to measure and predict our impact on the environment and future life on earth, our influence on future generations is greater and easier to evaluate than at any other point in human history. Developments such as climate change present us with unique challenges for moral decision making.

Most deontologists maintain that moral decision making should be informed by distinctions such as the DDA. In this paper, I have argued that the challenge faced by defenders of the DDA regarding complicated cases, such as cases of letting oneself do harm,

is more urgent, but also more complicated than has been acknowledged so far. This is because real-world cases are very complex, and because the applicability of the DDA to such complex cases is unclear. In the final section, I have discussed different ways in which defenders of the DDA could respond to this challenge, and tentatively proposed to classify them as instances of allowing harm. If my argument is correct, then deontologists will need a more fine-grained account of causal sequences and their moral relevance to account for the moral status of complicated cases, and ultimately real-world decision making.

## References

- Bennett, Jonathan. (1993) 'Negation and Abstention: Two Theories of Allowing'. *Ethics*, 104, 75–96.
- . (1995) *The Act Itself*. New York: Oxford University Press.
- Campos, Andre S. (2018) 'Intergenerational Justice Today'. *Philosophy Compass*, 13, e12477.
- Foot, Philippa. (1967) 'The Problem of Abortion and the Doctrine of the Double Effect'. *Oxford Review* 5, 5–15.
- Hanna, Jason. (2015a) 'Doing, Allowing, and the Moral Relevance of the Past'. *Journal of Moral Philosophy*, 12, 677–698.
- . 2015b. 'Enabling Harm, Doing Harm, and Undoing One's Own Behavior'. *Ethics*, 126, 68–90.
- Kamm, Frances. (2007) *Intricate Ethics: Rights, Responsibilities, and Permissible Harm*. New York: Oxford University Press.
- Persson, Ingmar. (2013) *From Morality to the End of Reason: An Essay on Rights, Reasons, and Responsibility*. Oxford: Oxford University Press.
- Quinn, Warren S. (1989) 'Actions, Intentions, and Consequences: The Doctrine of Doing and Allowing'. *The Philosophical Review*, 98, 287–312.
- Thomson, Judith J. (1985) 'The Trolley Problem'. *The Yale Law Journal*, 94, 1395–1415.
- Williams, Bernard. (1981) *Moral Luck: Philosophical Papers 1973-1980*. Cambridge: Cambridge University Press.
- Wolf, Susan. (2001) 'The Moral of Moral Luck'. *Philosophic Exchange*, 31.
- Woollard, Fiona. (2013) 'If This Is My Body...: A Defence of the Doctrine of Doing and Allowing'. *Pacific Philosophical Quarterly*, 94, 315–341.
- . (2014) Book Review: 'Persson, Ingmar. From Morality to the End of Reason: An Essay on Rights, Reasons, and Responsibility. Oxford: Oxford University Press, 2013. Pp. 336. \$55.00 (cloth)'. *Ethics*, 125, 272–276.
- . (2015) *Doing and Allowing Harm*. Oxford: Oxford University Press.
- Woollard, Fiona, and Howard-Snyder, Frances. (2016) 'Doing vs. Allowing Harm'. In Zalta, Edward N. (ed.) *The Stanford Encyclopedia of Philosophy*, Winter 2016. (Metaphysics Research Lab, Stanford University).  
<https://plato.stanford.edu/archives/win2016/entries/doing-allowing/>.