The Constraint Against Doing Harm and Long-Term Consequences

Charlotte Franziska Unruh
Faculty of Philosophy, University of Oxford, Oxford, United Kingdom
charlotte.unruh@philosophy.ox.ac.uk

Abstract

Many people hold the constraint against doing harm, the view that the reason against doing harm is stronger than the reason against merely allowing harm, everything else being equal. Mogensen and MacAskill (2021) have recently argued that when considering indirect long-term consequences of our everyday behavior, the constraint against doing harm faces a problem: it has the absurd implication that we should do as little as possible in our lives. In this paper, I explore the view that, for behavior that does not increase anyone’s ex ante risk of suffering harm, the reason against doing harm is not stronger than the reason against merely allowing harm, everything else being equal. I argue that this view is plausible, show how it solves the paralysis problem, and defend it against objections.

Keywords

Doing/allowing – Harm – Moral Constraints – Longtermism – Paralysis problem

1 Introduction

Many people think that the moral reason against doing harm is stronger than the moral reason against merely allowing harm, everything else being equal.1

1 I will sometimes omit the “everything else being equal” qualification for brevity and ease of expression.
Call this claim the *constraint against doing harm*.\(^2\) For an example of doing harm, think of Agent who pushes a rock downhill towards Victim. The rock crushes Victim to death. For an example of merely allowing (failing to prevent) harm, think of Agent who does not stop a rock that is rolling towards Victim. Again, the rock crushes Victim to death. The constraint against doing harm says that the reason against pushing the rock is stronger than the reason against allowing the rock to roll towards Victim.\(^3\)

In the discussion of the non-consequentialist constraint against doing harm, the focus lies on cases in which harm results directly and immediately from an action.\(^4\) This is unfortunate, because it remains unclear how the constraint applies to real-world cases, in which behavior might lead to indirect and unpredictable harm in the long term.

Mogensen and MacAskill (2021) and Nye (2014) have recently raised a problem for the constraint against doing harm. When considering remote (indirect and unpredictable) harmful effects of our behavior, the constraint against doing harm seems to have the absurd implication that we should do as little as possible in our lives. This is the *paralysis problem*.

In this paper, I argue that the scope of the constraint against doing harm is restricted in the following way. For behaviors that increase someone’s ex ante risk of suffering harm, the reason against doing harm is stronger than the reason against merely allowing harm, everything else being equal. However, for behaviors that do *not* increase anyone’s ex ante risk of suffering harm, the reason against doing harm is *not* stronger than the reason against merely allowing harm, everything else being equal.\(^5\) I call this view the *ex ante view*. I argue that the ex ante view solves the paralysis problem: because everyday actions do not increase anyone’s ex ante risk of remote harm, the paralysis problem does not get off the ground.

The paper proceeds as follows. In section 2, I explain the paralysis argument in more detail. In section 3, I present the ex ante view and use it to respond to the paralysis argument. In section 4, I discuss the objection that the ex ante view has implausible consequences, and in section 5, the objection that the ex ante view fails to acknowledge the moral relevance of the ex post perspective.

\(^2\) The claim is also known as the Doctrine of Doing and Allowing Harm (Quinn 1989; Woollard 2015). I call it the constraint against doing harm, as I focus on the special moral significance of doing harm.

\(^3\) These examples are adapted from Woollard (2015, 3).


\(^5\) I thank an anonymous referee for very helpful suggestions on the formulation of the ex ante view.
The upshot, I claim, is that non-consequentialists should accept the ex ante view. The ex ante view delineates the scope of the constraint against doing harm in a plausible way while avoiding the paralysis problem. In the following, by “reasons,” I mean subjective (i.e., evidence- or belief-relative) pro tanto moral reasons.6

2 The Paralysis Argument

Mogensen and MacAskill (2021) present their ‘paralysis argument’ as a reduc-tio of non-consequentialist principles.7 The argument starts from the assumption that everything that we do will have an infinite number of causal effects. Anything that we do could change how others behave, where they go, who they meet, when and with whom they have children, and what these children do. In the long run, these indirect effects of our actions are extremely likely to change who will be subject to harm, and when harm will occur.

Even mundane actions are likely to have such far-reaching consequences, because all our actions can be identity-affecting, i.e., affect the identity of future people (Parfit 1984, 361; see also Greaves 2016). For example, the decision to introduce a friend to another friend might lead to a relationship that would otherwise never have started, and a child might be conceived that would not have been conceived otherwise. In this case, the existence of a particular child, and the actions of this child, are indirect effects of the decision to introduce one’s friends. Mogensen and MacAskill (2021, 3) give an example of a similarly innocuous action: driving to the supermarket to buy milk. By driving to the supermarket, one will disrupt traffic ever so slightly. This disruption might have identity-affecting knock-on effects and cause people to exist who would not have existed otherwise – people who might later cause a fatal car accident. Mogensen and MacAskill point out that road accidents, even fatal ones, are common and conclude:

‘Somewhere down the line, therefore, we can expect that some people will die young who would otherwise have lived young and healthy lives

---

6 For the distinction between subjective and objective rightness, see Smith (2010).
7 Nye (2014) gives an argument towards similar conclusions based on chaos theory. However, Mogensen and MacAskill (2021, 2) suggest that chaos theory here is a ‘red herring’ as chaotic systems can still be relatively stable. To pre-empt such worries, I focus on Mogensen’s and MacAskill’s argument in the following.
if only you had not disturbed the traffic ever so slightly by driving to the supermarket to buy milk on that fateful day’

MOGENSEN AND MACASKILL 2021, 3.

I call cases such as the supermarket case *paralysis cases*. In paralysis cases, a mundane behavior is likely to lead to remote (i.e., indirect and unpredictable) harms in the future via a long chain of events. For our discussion, I assume that the behavior in question has no proximate\(^8\) harmful effects.\(^9\)

The paralysis argument proceeds as follows, as illustrated by the supermarket case. Mogensen and MacAskill assume that if a harm arises unpredictably as a result of driving to the supermarket, the agent counts as having done harm. Further, if a harm arises unpredictably as a result of sitting motionlessly at home, the agent counts as having merely allowed harm. Since the long-term consequences of actions are unpredictable, agents have no greater reason to expect that either behavior is more likely to bring about harm.

Given this, the constraint against doing harm gives the agent a stronger pro tanto reason against driving to the supermarket than against sitting motionlessly at home. This is because the constraint against doing harm gives agents a stronger reason against doing harm (by driving to the supermarket) than against merely allowing harm (by sitting motionlessly at home). Generalizing this point leads to the conclusion that the constraint against doing harm gives agents a reason against doing anything at all (Mogensen and MacAskill 2021, 6). This is absurd.

So far, I made a simplifying assumption, namely that harm will occur with certainty. However, in real-world contexts, we do not know with certainty whether harm will occur. What does the constraint against doing harm say about these cases? Mogensen and MacAskill (2021, 5), drawing on Lazar’s (2018; see also 2017) theory of deriving subjective from objective permissibility, answer this question by viewing decision situations as choices between options with different possible outcomes. Lazar proposes a standard of subjective permissibility which balances demands to maximize expected choiceworthiness with considerations of cost to the agent. The choiceworthiness of

---

\(^8\) I borrow the distinction between proximate and remote consequences from the legal literature on proximate causation (Moore 2019). While there is no universally accepted definition, proximate consequences tend to be direct and foreseeable, whereas remote consequences tend to be indirect and unforeseeable. I thank an anonymous referee for suggesting this distinction for defining paralysis cases.

\(^9\) For example, in the supermarket case, I ignore the risk of immediate accidents, or concerns about emissions and pollution.
an option can be influenced by moral constraints. For example, non-consequentialists will rank an outcome in which A has killed B as less choiceworthy than an outcome in which A has merely failed to save B (Lazar 2017, 585). On this theory, agents have a stronger subjective reason against behavior that is likely to constitute doing harm than against behavior that is likely to constitute merely allowing harm. Mogensen and MacAskill conclude (D denotes a doing behavior, A an allowing behavior, and O a morally relevant outcome):

‘[Y]ou have greater pro tanto reason to choose A rather than D in light of the possibility that O involves a harm to some individual. After all, you have greater objective reason not to do harm than to allow harm, and you have no more reason to expect harm to occur as part of O conditional on your choice of A rather than D.’

Mogensen and MacAskill 2021, 6.

This view seems prima facie plausible. Proponents of the constraint against doing harm think that there is a morally relevant difference between doing and merely allowing harm. It seems plausible that they should also think that there is a morally relevant difference between a behavior that is likely to constitute doing harm and a behavior that is (equally) likely to constitute merely allowing harm.

Anticipating objections to the paralysis argument, Mogensen and MacAskill discuss at length the possible responses on behalf of non-consequentialists and argue that these do not prove satisfactory. I agree with Mogensen and MacAskill that defending non-consequentialists from the paralysis problem is not an easy task. To illustrate, let me briefly explain why some apparently promising attempts to resolve the paralysis problem fail.

First, one might think that indirect consequences are too far removed from the agency of the driver in the supermarket example to still count as doing harm.10 However, insofar as it is foreseeable that there will be significant indirect consequences, it seems that they can give agents a reason to refrain from acting. (Consider a case in which pulling a lever will lead to many future harms. Even if the concrete instantiations remain unpredictable, you should not pull the lever.)

---

10 Mogensen and MacAskill (2021, 7–14) discuss and reject several lines of argument to the conclusion that everyday behavior that leads to indirect, unpredictable long-term harm does not constitute doing harm in the sense of the constraint. For example, they argue that we can do harm in a morally relevant sense, even when the harmful consequences are unlikely or occur via long causal chains.
Second, most actions seem just as likely to lead to indirect benefits as to harms. However, these benefits do not suffice to balance out the harmful indirect long-term effects of actions, since non-consequentialists commonly think that doing harm has more moral significance than doing good. So, a reason for doing good does not outweigh a reason against doing equivalent harm (Mogensen and MacAskill 2021, 4–5; Kagan 1989, 121–25).

Third, the fact that there is intervening agency involved in long-term harms does not absolve agents of responsibility. Mogensen and MacAskill (2021, 9) give an example of an agent who knowingly sells weaponry to a murderer. Surely, the constraint against doing harm should prohibit selling the weapon.

Fourth, non-consequentialists might reject the claim that sitting motionlessly at home in the supermarket case counts as merely allowing harm. Bennett (1995, 96–100) argues that remaining completely motionless can, sometimes, constitute a doing behavior. However, even if sitting completely still does not constitute allowing harm, the Paralysis Argument remains a challenge for non-consequentialists. After all, it seems plausible that non-consequentialists should still minimize the harm that they do (rather than allow), and options that constitute allowing harm are unlikely to allow significantly more freedom than sitting motionlessly (Mogensen and MacAskill 2021, 12–13).

In sum, the paralysis argument poses a challenge for non-consequentialists and highlights the need for spelling out criteria for the scope of the constraint against doing harm in the real world, where actions have many remote consequences. It has been noted in the literature that the special moral significance of doing harm might not apply to behaviors with very remote harmful effects. However, so far, non-consequentialists have not offered criteria for delineating the scope of the constraint.

3 The Ex Ante View and Solving Paralysis

The ex ante view says that, for behaviors that increase anyone's ex ante risk of suffering harm, the reason against doing harm is stronger than the reason against merely allowing harm, everything else being equal. However, for behaviors that do not increase anyone's ex ante risk of suffering harm, the reason against doing harm is not stronger than the reason against merely allowing harm.

Bennett (1995, 5–6) and Woollard (2015, 17–18) explicitly exclude cases in which the causal chain leading to harm runs through the voluntary will of others, or depends on intervening coincidences, from their discussion. However, they do not provide a general account of the scope of the constraint against doing harm.
harm, everything else being equal. In other words, the scope of the constraint against doing harm is restricted in at least one way: it does not apply to behavior that does not increase anyone’s ex ante risk of suffering harm. In this section, I argue that the ex ante view is prima facie plausible.

I take inspiration from the recent contractualist literature in formulating the ex ante view. A topic that is debated in this literature is the question when actions that impose a small risk of harms on large numbers of people are permissible. At issue is the strength of the moral claims of potential victims of risky behavior. In discussing this question, Frick (2015) distinguishes an ex ante and an ex post view. Very roughly, according to the ex ante view, the potential victims’ claims are discounted by the ex ante likelihood that they will be affected. According to the ex post view, the potential victims’ claims correspond to the likelihood that someone (as opposed to the individual victim) will be harmed (see Frick 2015, 185).

The ex ante view can provide a plausible criterion for limiting the scope of the constraint against doing harm. In standard cases of doing harm, the agent increases the risk that the victim suffers harm, and so the condition described by the ex ante view is fulfilled. For example, when Ann punches Bob, she makes it more likely that Bob will experience pain. However, as I will argue, paralysis cases are non-standard cases. In paralysis cases, the agent’s behavior does not increase anyone’s ex ante risk of suffering harm. So, according to the ex ante view, the reason against doing harm is not stronger than the reason against merely allowing harm, and the paralysis problem does not get off the ground.

3.1 The Ex Ante View

Let me begin by introducing the ex ante view more carefully. Consider the following cases.

(Wheel Doing) A Giant Wheel of Fortune has the names of a million people on it. The wheel is not spinning. You can spin the wheel. If you spin the wheel, then the person whose name is selected will be killed.

(Wheel Allowing) A Giant Wheel of Fortune has the names of a million people on it. The wheel is spinning. The person whose name is selected

---

12 I assume that the number of people who might suffer harm is finite and ignore problems that arise in cases involving an infinite number of people. Thanks to an anonymous reviewer for prompting me to clarify this.
will be killed. You can stop the wheel. If you stop the wheel, no name will be selected, and nobody will be killed.\(^{13}\)

I assume that if you spin the wheel in (Wheel Doing), then your behavior constitutes doing harm. I also assume that if you fail to stop the wheel in (Wheel Allowing), your behavior constitutes merely allowing harm.\(^ {14}\) Spinning the wheel in (Wheel Doing) results in higher ex ante likelihood of each of the one million people being killed. Failing to stop the wheel in (Wheel Allowing) results in higher ex ante likelihood of each of the one million people being killed. The reason against spinning the wheel is stronger than the reason against failing to stop the wheel. Now consider

(Wheel Extra Spin Doing) A Giant Wheel of Fortune has the names of a million people on it. The wheel is spinning. You cannot stop the wheel. You can give the wheel an Extra Spin. The Extra Spin does not change the fact that one of the million people will be killed. The Extra Spin does also not change the risk of being selected for any of the million people. However, the following is likely to be true: Had you not given the wheel an Extra Spin, a different person would have been killed.\(^ {15}\)

Like spinning the wheel in (Wheel Doing), the Extra Spin in (Wheel Extra Spin Doing) is certain to result in the killing of one person. However, in terms of ex ante risk, the cases differ. In (Wheel Doing), the agent’s behavior slightly increases the ex ante risk of harm for each of the million people. In (Wheel Extra Spin Doing), the agent’s behavior does not increase the ex ante risk of harm for anyone. This is because the agent’s behavior does not increase the likelihood, for anyone, that they will be selected. Now consider

(Wheel-Extra Spin Allowing) A Giant Wheel of Fortune has the names of a million people on it. The wheel is spinning. You cannot stop the wheel.

\(^{13}\) These cases are inspired by a series of cases in Otsuka (2015) which also feature wheels.

\(^{14}\) Here, I follow Mogensen’s and MacAskill’s assumption (see Mogensen and MacAskill 2021, 5) that actions that lead to remote harm, such as driving to the supermarket generally count as doing harm, and omissions that lead to remote harm, such as sitting at home generally count as merely allowing harm. It might be argued that these cases do not count as doing or allowing harm in the first place (see the objections surveyed in section 2). If such an argument succeeds, it would also undermine the paralysis problem.

\(^{15}\) Woollard’s Extra Push case (Woollard 2015, 24), despite its name, is different from the wheel cases, as in her case the extra push is necessary for harm to occur.
A sudden wind gust is about to give the wheel an Extra Spin. You can now shield the wheel from the wind gust but fail to do so. The Extra Spin would not change the fact that one of the million people will be killed. The Extra Spin would also not change the risk of being selected for any of the million people. However, the following is very likely to be true: Had you shielded the wheel from the wind gust, a different person would have been killed.

In (Wheel-Extra Spin Allowing), like in (Wheel-Extra Spin Doing), your behavior does not increase the ex ante risk of harm for anyone. It does not change the likelihood, for any of the million people, that they will be selected.

According to the ex ante view, the reason against pushing the wheel in (Wheel-Extra Spin Doing) is not stronger than the reason against failing to shield the wheel in (Wheel-Extra Spin Allowing). (Note that this does not amount to the claim that the cases are overall morally equivalent, or that there is no moral reason against giving the wheel an extra spin.)

3.2 Defending the Ex Ante View

As mentioned above, the constraint against doing harm is usually illustrated and defended with reference to standard cases of doing and allowing harm. In standard cases, the agent interferes with the victim, making a difference to what happens to the victim. The agent’s behavior determines whether someone will be harmed. Intuitively, giving the wheel an extra spin lacks the special moral significance of doing harm in standard cases. I suggest that this is because the victim’s ex ante prospects remain unchanged. From the victim’s point of view, the agent does not make it any more likely that the victim suffers harm.

The constraint against doing harm is often thought to offer protection from harmful interference or imposition (Foot 2002, 82; Kagan 1989, 100; Woollard 2015, 102). Woollard (2015) offers a comprehensive account of the distinction between doing and allowing harm and a defense of the moral relevance of the distinction. According to Woollard, the constraint against doing harm protects agents from ‘causal imposition’ (Woollard 2015, 99). Roughly, when agents do harm, their behavior is part of a causal chain leading to an effect on the victim. When agents allow harm, the causal chain from their behavior does not reach...

---

16 For example, the agent’s behavior might be morally problematic because the agent is complicit in a wrongdoing or acts with bad intentions.
the victim, and the behavior is better described as a background condition of the harm (Woollard 2015, 99–105).

Given this justification of the constraint against doing harm, we can see a difference between standard cases of doing harm and cases of doing harm without changing the ex ante likelihood of harm. In (Wheel Doing), if the agent spins the wheel, the agent changes the situation for the potential victims, by imposing a threat of harm on them. In contrast, in (Wheel Extra Spin Doing), if the agent gives the wheel an extra spin, the agent does not change the situation for the potential victims. The causal chain that leads to harm is already underway: the wheel is spinning. The extra spin does not change anyone’s ex ante likelihood of suffering harm.

Drawing on this observation, I suggest that in (Wheel Extra Spin Doing), the agent does not impose on the potential victims in a relevant way. Here, I draw on a remark by Woollard, who points out that a behavior leading to a harmful effect might not be seen as an imposition if it is ‘mediated by, for example, running through another person’s will or depending on a series of unforeseeable coincidences’ (Woollard 2015, 101). Woollard does not elaborate on which conditions must obtain for an effect to be ‘mediated’ in this way. Intuitively, however, intervening agency and coincidences mediate the effect on the victim because they are external to the agent and lower the agent’s control over whether harm occurs. In (Wheel Extra Spin Doing), the fact that the wheel is already spinning constitutes a threat to the victim that is external to the agent. The agent neither diminishes nor magnifies the original threat by giving the wheel an extra spin. The agent does not (and cannot) change anyone’s ex ante risk of suffering a harm. Like in the cases described by Woollard, I suggest that

---

17 A related point has been noted in the literature regarding pre-emption cases, in which Victim will certainly be harmed no matter what Agent does. In discussing these cases, Woollard states that harming someone is ‘significantly easier to justify if the harmed person is not made worse off’ (2012, 688). For example, it seems easier to justify killing someone to save a life, if the person will be killed regardless. Similarly, Gardner (2017, 86) introduces a ‘Redundant Harming’ principle to explain the strength of the reason against harming. According to Gardner, ‘an action redundantly harms an individual when it causes her to suffer a harm that she would have suffered anyway, had the action not been performed.’ The principle then says that redundance weakens the reason against harming (Gardner 2017, 86). For similar reasons, it seems plausible that the reason against doing and allowing harm in the Extra Spin cases is weaker than in the regular Wheel cases. This line of argument, if successful, further weakens the paralysis argument.
in the extra spin case, the agent’s behavior does not constitute an imposition in the relevant sense.18

One might object that the agent imposes on the victims by influencing who will be harmed.19 In some cases, changing merely who will be harmed seems to matter morally. Consider the phenomenon called ‘moral inertia’ by Sartorio: ‘moral pressure to leave things “as is,” in other words, to fail to intervene’ (Sartorio 2008, 118). Sartorio illustrates the phenomenon with the case of a trolley that is heading towards one person, but can be diverted to a side track, where one other person is trapped (2008, 119–120). It seems prima facie wrong to divert the trolley. Sartorio explains this intuition by saying that the victim on the side track has a complaint against the agent: if the agent turns the trolley, the victim faces a threat that the victim did not face before (Sartorio 2008, 120).20 However, in (Wheel Extra Spin Doing), the potential victims cannot make this complaint. After all, nobody faces a threat that they did not face before the agent’s action. So, moral inertia does not arise.

In sum, the justification of the constraint against doing harm refers to standard cases of doing and allowing harm in which the agent imposes on the victim by making a difference to whether a harm occurs. However, in the extra spin cases, the agent does not change anyone’s ex ante risk of suffering harm. I argued that the agent’s behavior does not constitute an imposition on the victim in the relevant sense. If this is correct, then we have reason to think that

---

18 One might object that in the case of intervening agency, it matters that the intervening agent acts after the primary agent. In contrast, the wheel is set in motion before the agent acts in the extra spin cases. I think that what matters is not the temporal order of events but whether and how the agent influences the harmful effect. However, to address this worry, let me offer a different version of the Wheel Extra Spin Doing case. In this case, the wheel is not spinning. The agent gives the wheel an extra spin. Nothing happens. However, later the wheel is set in motion. One person is selected and killed. Had the agent not given the wheel an extra spin, then it is likely that a different person would have been selected. I believe that my argument still succeeds using this revised version of the case.

19 Influencing who will be harmed does not always seem morally problematic. For example, consider the following case, given by Hanser (1990, 62): an agent can decide whether to adopt a new highway plan. The number of future traffic accidents will be the same either way, but the accidents will likely involve different people if a different highway plan is adopted. It seems permissible to adopt the new highway plan.

20 In a different discussion, Sartorio argues that diverting threats towards the lesser number might be easier to justify than initiating new threats. In the diversion cases, agents might be the cause of a particular death but not the fact that someone dies (Sartorio 2011, 446–48). As suggested to me by an anonymous referee, the agent in the Extra Spin cases might also not be the cause of the fact that someone dies. It seems plausible that this causal difference makes a moral difference. If this is right, it provides a reason to accept the view that the reason against doing harm is weakened in the Extra Spin cases.
the constraint against doing harm does not apply to behavior that does not increase anyone’s ex ante risk of suffering harm, everything else being equal.

3.3 Paralysis Cases as Extra Spin Cases

To complete my argument, I now argue that paralysis cases are like Extra Spin cases, not like standard cases of doing harm. In paralysis cases, agents do not increase anyone’s ex ante likelihood of (remote) harm. It is plausible to assume, for every possible remote harm, that it is just as likely that an everyday action causes this harm as it is that the action prevents this harm from occurring (see Mogensen and MacAskill 2021, 3). To see this, consider a future person who lives 200 years from now. Is it more likely that driving to the supermarket today causes this person to die prematurely? Or is it more likely that driving to the supermarket today prevents this person from dying prematurely? Given that the long-term effects of driving to the supermarket are unpredictable, it seems reasonable to assume that both are equally likely.

Support for this reasoning can be derived from an argument made by Greaves (2016). Greaves classifies cases in which we are ‘clueless’ about the future consequences of actions in simple and complex cases (complex cases will not interest us here). In simple cases, we consider unforeseeable effects: ‘while they could result from (say) some particular Act A, they could equally easily, and in precisely analogous ways, result from any of the relevant alternative acts.’ This is why the indifference assumption, roughly the assumption that it is equally likely that an outcome arises from each of the relevant alternative acts, is plausible. Here is how Greaves illustrates simple cases with the example of an everyday behavior:

‘While there are countless possible causal stories about how helping an old lady cross the road might lead to (for instance) the existence of an additional murderous dictator in the twenty-second century, any such story will have a precise counterpart, precisely as plausible as the original, according to which refraining from helping the old lady turns out to have the consequence in question and it is intuitively clear that one ought to have equal credences in such precise-counterpart stories.’

Greaves 2016, 322.

I thank an anonymous referee for prompting me to make these arguments explicit, and for pointing me towards Greaves’ article.
I think that paralysis cases are relevantly like such simple cases. There are many possible ways in which driving to the supermarket today might lead to a fatal accident in 100 years. Similarly, there are many equally possible ways in which failing to drive to the supermarket today might lead to a fatal accident in 100 years (such that driving to the supermarket could have prevented the accident). Imagine a scenario in which Bob is run over by a car in 2121. It seems reasonable to assume that it is just as likely that Bob dies as a result of my driving to the supermarket in 2021 as that Bob dies as a result of my failing to drive to the supermarket in 2021. Driving to the supermarket is thus analogous to giving the wheel an extra spin in (Wheel-Extra Spin Doing). Similarly, it seems reasonable to assume that it is just as likely that Bob is run over as a result of my staying at home as that Bob is run over as a result of my failing to stay at home. Staying at home is thus analogous to failing to shield the wheel in (Wheel-Extra Spin Allowing).22

I argued that the reason against (Wheel-Extra Spin Doing) is not stronger than the reason against (Wheel-Extra Spin Allowing), everything else being equal. Similarly, the reason against driving to the supermarket is not stronger than the reason against staying at home. If the reason against driving to the supermarket is not stronger than the reason against staying at home, then a crucial premise for the paralysis argument is undermined. The argument, then, does not get off the ground.

4 The Ambulance Objection and Pro Tanto Harm

Mogensen and MacAskill (2021, 9–10), following Kamm (2007, 274–75), argue that defenders of the constraint against doing harm should refrain from

22 There is reason to suspect further complications. The same behavior can both increase and decrease the risk of a particular future harm. This is because every behavior has multiple causal effects. For example, a behavior might influence the identity of both the father and the mother of a harm doer. One effect of a behavior might also prevent another effect from causing harm. For example, a behavior might influence the identity of a parent who takes their child out on a boat ride and the lifeguard who saves the child from drowning. This might suggest that the effects of our actions might be less vast than assumed. Moreover, wheel cases are one-off decision scenarios, but we face many small decisions in life every single day. This might give rise to further relevant moral considerations, such as rights to impose risks on other people in order to live our own lives (see, e.g., Steigleder’s argument for rights to impose minor risks (Steigleder 2016) and Woollard’s concept of ‘normative imposition’ (Woollard 2015; 2013)).
evaluating outcomes from an ex ante perspective. This is because such a view would have implausible implications in the following case:

Ambulance: ‘A community considers the purchase of an ambulance. The ambulance has an on-board artificial intelligence and is rigged with special brakes. If the ambulance is hurrying to the hospital, the on-board AI will kick in and make it impossible to stop the ambulance from running over someone in its way if this is necessary to save a greater number of people being transported on-board.’

Mogensen and MacAskill 2021, 10; the case is due to Kamm 1996, 2:303.

The decision to buy the ambulance is intuitively impermissible. However, it seems that ex ante reasoning does not support this judgment. After all, buying the ambulance does not increase the ex ante risk of harm for anyone in the community. (The decision is more likely to save the life of any community member than to kill them.) As Mogensen and MacAskill (2021, 10) note, the reason why buying the ambulance is impermissible seems to be the same kind of harm-based reason that the constraint against doing harm generates. Intuitively, it is wrong to kill one person, even if this produces more good overall.

However, non-consequentialists can explain why buying the ambulance is impermissible. The key idea is that the constraint is not a principle about overall harm; rather, it is about pro tanto harm. This is because compensated harm is still harm. The constraint should tell us to refrain from doing harm, even when we compensate. Consider, for example, the company that sells products that increase the risk of cancer A, but reduce the risk of cancer B. Even if these equal out such that the expected overall impact on health is 0, the company still increases the risk of cancer A, and therefore the expected pro tanto harm of suffering this particular cancer.

The suggestion, therefore, is not that the forward-looking, action-guiding constraint against doing harm only applies to actions that worsen anyone’s ex ante likelihood of suffering harm overall. Rather, the constraint against doing harm applies to actions that increase anyone’s ex ante likelihood of suffering pro tanto harm.

One might object that the people in (Ambulance) are harmed in the same pro tanto respect. Some people die in traffic accidents caused by cars, others in traffic accidents caused by ambulances: this looks like the same type of harm. However, it is not the same type of harm. The decision to buy ambulances does not reduce the number of car crashes. It ensures that those who are involved in car crashes are more likely to get help. One might describe what goes on in
the ambulance case as follows. The decision makes it more likely that people will be involved in car accidents. And importantly, it makes it more likely that people will receive help if they get involved in car accidents. Consider this variant of the case:

(Ambulance*) A community considers the purchase of an additional ambulance. Due to factors such as weather conditions and poor driving skills, ambulances regularly cause accidents in this community. It is expected that an additional ambulance would increase the risk for members of the community to be involved in accidents caused by ambulances run by the community. The community also plans to run extra driver’s training for ambulances. Such training would decrease the risk for members of the community to be involved in accidents caused by ambulances run by the community.

In this case, the constraint against doing harm intuitively does not give the community reason to go against their plans. This is because the plans, overall, do not make it more (ex ante) likely for anyone affected that they will suffer harm in any respect, at any time.²³

5 The Objection from the Ex Post View and Being a Harm-Doer

Here is another objection to the ex ante view. Even if behavior in paralysis cases does not result in higher ex ante risk of harm for anyone, agents in paralysis cases make it the case that they will have done harm. There will be a victim that is harmed by the agent that would not have been harmed, had the agent acted differently. Intuitively, agents should care about not being harm doers.

²³ Further justification for this judgment might be drawn from the contractualist literature on the question when an action is justifiable to an agent. Frick’s decomposition test, ‘according to which an action is justifiable if and only if the actions it licenses us to perform are justifiable to each person at each temporal stage’ (Frick 2015, 181) seems to imply that buying the ambulance in (Ambulance) is not justifiable, since it does not seem justifiable to the pedestrian when the on-board AI kicks in. On Gordon-Solmon’s proposed alternative to the decomposition test, deliberators in contractualist settings compare their claims under ignorance of their identity but not their group memberships (Gordon-Solmon 2019, 273). On this view, buying the Ambulance in (Ambulance) might not be justifiable to the pedestrian, if she knew that she would (likely) be a pedestrian and not a patient. In contrast, in (Ambulance*), there is no point in time at which buying the ambulance is not justifiable to pedestrians.
However, it seems that the ex ante view cannot explain this intuition. What this shows, the objection goes, is that the ex post perspective matters morally. The reason against a behavior that is likely to constitute doing harm is stronger than the reason against a behavior that is likely to constitute allowing harm, if someone will suffer harm as a result of the behavior.\(^{24}\) I call this the *ex post view*.\(^{25}\) The ex post view does not exclude cases like the Extra Spin cases from the scope of the constraint against doing harm.

According to the ex post view, there is a stronger reason against spinning the wheel in (Wheel Extra Spin Doing) than against failing to shield the wheel in (Wheel Extra Spin Allowing). This is because it matters whether we will have done harm or merely allowed harm, regardless of whether we have increased the odds of harm occurring.

My response is that the burden of proof is on the objector to show that the ex post perspective matters per se. This is because the constraint against doing harm is, in the first instance, a principle about paradigm cases of doing and allowing harm. Recall the case from the introduction in which Agent pushes a rock towards Victim. From an ex ante perspective, Agent makes it more likely that Victim will suffer harm, and from an ex post perspective, Agent has done harm to Victim. But paralysis cases are different from these paradigm cases: agents do not increase ex ante risks. Given this, defenders of the ex ante view are justified in shifting the burden of proof onto their opponents. The critic of the constraint needs to explain why the constraint should retain (much of) its original strength in paralysis cases, even though these cases are unlike paradigm cases.

In response, defenders of the ex post views might point to the strength of the intuition that doing something that makes someone suffer harm is bad. Imagine a case in which, if Agent does not shoot Victim, then Villain will shoot Victim. Agent does not worsen Victim's ex ante prospects by shooting Victim, since Victim will be shot either way. However, Agent clearly should not

---

\(^{24}\) This objection is inspired by the view defended by Otsuka (2015, 84) that we should discount complaints against being harmed by the likelihood that someone will be harmed. (More commonly, it is argued that we should discount complaints by the likelihood that an identified individual will be harmed (see, e.g., Kumar 1999, 295; Reibetanz 1998, 301)).

\(^{25}\) This does not rule out a hybrid view, according to which actions that are likely to do harm (ex ante) and actions that have resulted in harm (ex post) are both constrained. I deliberately leave it open whether the defender of the ex post view could adopt such a hybrid view. However, there is reason to think that such hybrid views might give counterintuitive results in the case of asymmetric information (Reibetanz 1998, 302; Fried 2012a, 54).
shoot! Do we not, the objection might go, need the ex post view to explain this judgment?

I do not think so. It is plausible to think that ex ante and ex post perspectives operate in different domains. Prohibitions of actions are grounded in ex ante evaluation, whereas compensation for outcomes is grounded in ex post evaluation (Fried 2012b, 237–38). These perspectives are importantly distinct. To illustrate, Ann, whose speeding has caused an accident, is morally responsible and legally liable, even if Bert’s speeding at the same time was similarly likely to cause the accident. However, it would be a mistake to infer from this that before the accident occurred, Ann had a stronger reason to refrain from speeding than Bert.

“Wait!” the objector might say. “You have just admitted that a person who suffers harm as a result of giving the wheel an extra spin in (Wheel-Extra Spin Doing) has a valid compensation claim against the agent. However, the agent knows that such claims will arise, simply because it is certain that some slot will be selected after they have spun the wheel. The agent should acknowledge these claims now – before the harmful outcome has occurred. And surely, this gives the agent a reason against acting now!”

I agree that the agent should acknowledge the fact that her behavior might give rise to claims of those who will suffer harm as a result of her action. However, it is a mistake to think that this necessarily gives the agent a stronger reason against acting than against refraining from acting. This is not only because compensation claims seem misdirected when raised against agents who have contributed to an outcome (very) indirectly and unforeseeably (as opposed to those who have contributed directly and foreseeably). Rather, future compensation claims can be acknowledged by taking precautions. Reasonable precautions seldom require abandoning the risky activity altogether, especially if doing so would be costly. Reasonable precautions can consist in minimizing the risk that actions impose on others, e.g., by providing safety equipment for construction workers or organizing extra driver’s training in (Ambulance*). Another way of taking precautions might be setting up systems or institutions that can mitigate harm or compensate those who end up suffering as a result of one’s actions, e.g., by setting up funds or insurance systems for workplace accidents.

Now, one might think that agents in paralysis cases automatically take such precautions. Agents in paralysis cases do not increase anyone’s ex ante risk of harm. They are just as likely to benefit any given person as they are to harm them. Admittedly, the actual benefits are unlikely to occur to those who will be harmed in a way that compensates them. However, agents in paralysis cases cannot effectively target their compensation efforts, since they have no idea
which harms their actions will cause to whom and when. Moreover, any compensatory measure that agents would put in place would give rise to the same risks of harm as the risk-imposing actions themselves. For these reasons, the fact that one’s everyday actions are likely to result in harm in the future does not automatically give rise to clear duties of precaution or compensation.26

6 Conclusion

I argued that non-consequentialists should accept the ex ante view on the constraint against doing harm, which says that the reason against doing harm is stronger than the reason against merely allowing harm only for behaviors that increase an individual’s ex ante risk of suffering harm. If this is right, the paralysis problem does not get off the ground. This is because everyday behaviors do not increase anyone’s ex ante likelihood of suffering remote harm. I further argued that the fact that what we do will have many effects in the world, good and bad, does not provide moral reasons against doing things. However, if the intuition behind the ex post view is correct, then we are more responsible for the results of our doings than for the results of our allowings. I argued that even if this is right, it would be a mistake to infer that we should stop doing things to avoid responsibility for future harms. Taking responsibility might instead require us to care more about the future that we cause to exist.

26 Finally, here is another apparent counterexample to the ex ante view. Imagine that you can save a stranger by running a trolley down a track. One of a thousand people will be tied to the track (it is uncertain which individual that will be). It might seem that the ex ante view should permit running over the person: a 1/1000 chance of death is much better than the certain death of the stranger, whereas the ex post view has the intuitively correct implication that we should not run over the stranger. However, the ex ante view does not say anything about how to aggregate. The ex ante view is a claim about the comparative strength of reasons against doing and allowing harm. It says that there is a stronger reason against running the trolley down the track, thereby increasing the risk of harm slightly for each of the 1000, than against failing to stop a trolley, thereby failing to decrease the risk of harm slightly for each of the 1000, all else being equal. While I agree that aggregation remains an important and unresolved issue in the debate on cases of doing harm under risk, for the purpose of my discussion, I remain neutral on the debate. Thanks to an anonymous reviewer for providing this case, and for pressing me to think about these cases and clarify my argument.
Acknowledgements

I am grateful to anonymous reviewers of this journal for extremely helpful comments and questions. For excellent discussion and feedback, I am thankful to the audiences at the Global Priorities Institute Workshop, University of Oxford, December 2020, and the Research Colloquium at the Chair for Practical Philosophy, Humboldt-University Berlin, December 2020.

Funding

This paper draws on material first developed in chapter 6 of Charlotte Franziska Unruh. 2020. “Doing and Allowing Harm to Future Generations.” PhD diss., Philosophy Department, University of Southampton, which was funded by an Arts and Humanities Research Council Grant (AH/L/503939/1).

Biographical Note

Charlotte Franziska Unruh is an Early Career Research Fellow specializing in moral and political philosophy at the Institute for Ethics in AI, Faculty of Philosophy, University of Oxford.

References


