The paralysis argument

William MacAskill, Andreas Mogensen

Global Priorities Institute | September 2019

GPI Working Paper No. 6-2019

The Paralysis Argument

ABSTRACT: Given plausible assumptions about the long-run impact of our everyday actions, we show that standard non-consequentialist constraints on doing harm entail that we should try to do as little as possible in our lives. We call this the Paralysis Argument. After laying out the argument, we consider and respond to a number of objections. We then suggest what we believe is the most promising response: to accept, in practice, a highly demanding morality of beneficence with a long-term focus.

1.

Some effects of our actions are immediate and predictable. Many more are delayed and unpredictable. It's plausible that the long-run moral importance of anything you do is determined primarily by these indirect effects: far more people are harmed or benefited in ways you can't possibly foresee than are helped or hindered foreseeably as a result of what you do. Lenman (2000) argues that this poses a significant problem for consequentialism, because we are left clueless about the value of our actions. In this article, we argue that the problem for non-consequentialists is greater still.

Our argument is as follows. Let's assume, as above, that any one of your actions will generate many knock-on effects, spreading onward through time and magnifying in significance as history unfolds. You can't foresee the ways in which harm might arise indirectly as a result of doing one thing as opposed to another. Nor can you foresee the ways in which behaving this way as opposed to that might unexpectedly benefit people later on. According to most non-consequentialists, reasons against doing harm are weightier than reasons to benefit. Since you have no greater reason to expect that benefits as opposed to harms will predominate among the indirect effects of any action you perform, it therefore

seems that you should try as best you can to avoid bringing about any significant indirect effects through your actions at all. Since virtually anything you do will inevitably result in significant numbers of indirect harms, you should therefore try to do as little as possible.

We call this the *Paralysis Argument*. Its conclusion isn't without precedent in the history of philosophy. The most devout followers of Jainism take extraordinary precautions to avoid harming any living thing, and the ideal death in Jain ethics is by *sallekhana* ('thinning out'), which involves sitting motionless until you starve to death, ensuring through this ultimate act of passivity that you harm no living being (Webb 2018). But we don't use the Paralysis Argument to support the practice of *sallekhana*. The point is to show that endorsing the standard non-consequentialist asymmetries between doing versus allowing and harming versus benefiting leads to absurd conclusions when we take account of the long-run effects of our actions.[1] The aim of this paper is to set out the Paralysis Argument in its most convincing form and consider how non-consequentialists might respond to it. We argue that the most promising response involves accepting what in practice amounts to a highly demanding morality of beneficence, with a long-term focus. The Paralysis Argument may therefore reveal a tension between the twin pillars of common-sense deontology: belief in deontological constraints and belief in a modestly demanding morality of beneficence (Kagan 1989; Scheffler 1982).

[1]In this respect, our use of the Paralysis Argument resembles a recent paper by Nye (2014), which draws similar conclusions from chaos theory. We actually think chaos theory is a red herring in this debate. Sensitive dependence on initial conditions is not unique to chaotic systems. Furthermore, the 'strange attractors' associated with chaotic systems mitigate the significance of sensitive dependence. Chaotic systems are characterised by micro-disorder within macro-order (Smith 1998), in light of which they can actually be more robust to small perturbations than linear systems (Gleick 1987: 292-3).

Here is the plan. In section 2, we offer a more careful and rigorous statement of the Paralysis Argument, beginning with a survey of its descriptive assumptions and then exploring their moral implications. In section 3, we discuss six *prima facie* plausible objections to the argument that we think have relatively little *ultima facie* plausibility: (i) that the indirect and unforeseeable consequences of our actions are morally irrelevant; (ii) that you do not count as doing harm if the causal chain goes through the voluntary acts of other agents; (iii) that some intermediary event linking present behaviour to future unforeseeable harm fails to count as *substantial* in the sense outlined by Woollard (2015); (iv) that any harms we cause indirectly and unforeseeably are justifiable by appeal to considerations of *ex ante* Pareto superiority; (v) that the argument is undermined by the Non-Identity Problem; and, finally, (vi) that the argument rests on a mistaken conception of what it means to allow some outcome to come about. In section 4, we consider a response that hinges on the observation that the Paralysis Argument ignores the possibility of actions aimed at improving the very long-term future. Building on this observation, we argue that there is a way to escape paralysis, so long as we endorse a highly demanding conception of beneficence with a long-term focus. Section 5 offers a summary and conclusion.

2.

The aim of this section is to offer a more careful and rigorous formulation of the Paralysis Argument. There are lots of objections to the argument that we'll gloss over for the time being, so don't be alarmed if some steps in the argument seem to overlook complications. We'll address what we see as the key objections in sections 3.

We begin by examining the descriptive assumption on which the argument rests: namely, that our actions generate many unpredictable indirect effects, magnifying in

significance over time. Why should we believe this? The primary reason is that the facts of reproductive biology intersect with the metaphysics of transworld identity in such a way as to entail the ubiquity of *identity-affecting actions* (Greaves 2016; Lenman 2000). By an 'identity-affecting' action, we mean an action whose performance affects which persons belong to the total population of everyone who ever exists.

We take Parfit (1984) to have shown that our existence is remarkably fragile across nearby possible worlds: if your parents had not conceived a child at around the time at which they did in fact conceive, then you would not exist. Almost anything we do can be expected to affect the timing of some reproductive event and thereby change which people come to exist. For example, Greaves (2016) highlights that any act impacting the local traffic will ever so slightly quicken or slow the journeys of countless people. Some of these people were going to conceive a child on the day in question. Since a difference of just a few milliseconds is likely to affect which particular sperm out of the 200 million sperm in a typical ejaculation fertilises the egg, any action that causes such a delay for sufficiently many people is likely to be identity-affecting, ensuring that someone is born who would otherwise not have existed.[2]

Note, furthermore, that this action will be causally responsible for everything this person does, for good or ill. This includes the person's effects on later reproductive events, ensuring that the identity-affecting consequences of your actions grow quickly within the

[2] Assuming that on average a person produces one child in their lifetime, if your action affects, in any way, the timeline of 700,000 days, then you have probably altered a conception event. This number is high enough that probably not every drive to the supermarket is identity-affecting; but such a drive still creates a significant risk of being identity-affecting, which is sufficient for the purposes of our argument. It's also plausible that the slight changes in another person's day will slightly impact subsequent days of theirs; if by driving to the supermarket you (ever so slightly) affect the nature of a person's whole lifetime, then, on average, you will have altered the timing of a conception event.

foreseeable future. More and more reproductive events will be altered. More and more people will be born who would not have existed but for your action. Everything they do will be indirectly traceable to your act. Very many of these consequences will be harmful or beneficial to other people. For example, some of these people will inevitably get into car wrecks that would not otherwise have occurred. Over the course of a typical lifetime, the average motorist in the US will be involved in three to four road accidents, and three out of every 1,000 accidents involves fatalities (Toups 2011). Somewhere down the line, therefore, we can expect that some people will die young who would otherwise have lived long and healthy lives if only you had not disturbed the traffic ever so slightly by driving to the supermarket to buy milk on that fateful day. But we arguably have equally good reason to expect that some accidents will be averted that would otherwise have occurred. Somewhere down the line, we can expect that someone's life will be saved that would otherwise have been ended by a driver suffering a momentary lapse of attention - a driver who happens not to exist because of you. With approximately 100 people in the US dying every day in motor vehicle crashes, as time runs on and on, the number of fatal accidents caused and averted by your actions may be expected to grow and grow, running into the hundreds, the thousands, perhaps even the millions.

Granting this, what follows, morally speaking? On standard non-consequentialist views, the implication seems to be that we ought to try to do nothing at all. To see this, let's first consider the following thought experiment.

The Dice of Fortuna

There exists a pair of dice of ancient and mysterious provenance, called 'The Dice of Fortuna,' in honour of the Roman goddess of chance. The dice are sealed in a box. You don't know how many

sides each die has, nor who made them, nor how they were fashioned, nor from what. You do know that their faces are numbered. You also know that the dice have been enchanted in such a way that if you choose to roll the dice and the result is below the average value of the sum of the numbered faces on both dice, this will save a life. However, if you get a number equal to or greater than the average, someone will be killed. You will gain $1 if you shake the box, rolling the dice. Ought you to do so?

It seems clear to us that, on standard non-consequentialist views, you ought not to roll the dice. The chance of saving a life does not morally offset the chance of killing someone, and nor does the guarantee of gaining $1. But *The Dice of Fortuna* is closely analogous to the situation in which we find ourselves on a day-to-day basis. For reasons we have just discussed, when we drive to the supermarket we have an enormous effect on others, both causing deaths and saving lives. On standard non-consequentialist views, the lives saved do not straightforwardly compensate for the deaths caused, and neither do the benefits of buying groceries. So the non-consequentialist should regard driving to the supermarket, and almost any other identity-affecting action, as immoral.

We believe we reach the same conclusion if we rely on standard theoretical statements of the non-consequentialist asymmetries between doing and allowing and harming versus benefiting. In order to explain why this is so, we'll start with a few comments on these asymmetries.

First and foremost, we have the well-known *Doctrine of Doing and Allowing* (DDA), according to which your reasons against doing harm are stronger than those against allowing otherwise equivalent harms to befall people (Woollard and Howard-Snyder 2016). This gives us one important piece of the puzzle. But we're convinced by Kagan (1989: 121-5) that the DDA is insufficient if we want to fully capture those intuitions standardly elicited in support of constraints on doing harm. Consider the following case, due to Foot (1984):

Rescue:

You are hurrying to the beach in your jeep in order to rescue five people whose lives are threatened by the incoming tide. The quickest way to the beach involves taking a narrow and rocky path. Unfortunately, a person is walking on the path, and can't be expected to get out of your way in time. No other path will get you to the beach on time. In order to save the five, you must drive over the one on the path, killing her.

Most people believe it would be wrong to continue along the path. This might be explained by the idea that doing harm is harder to justify than letting an otherwise equivalent harm befall someone. This allows us to say that it's not as bad to allow the five people to die as to kill the one on the path. However, Kagan notes that in cases like *Rescue*, the course of action that involves doing harm also involves actively providing benefit to other people: namely, the five who are threatened. The fact that doing harm is worse than allowing harm fails to justify foregoing this action unless something is also said about how doing harm stands to doing good.

To capture our intuitions, Kagan argues, we must also assume (what we'll call) the *Harm Benefit Asymmetry* (HBA), according to which reasons against doing harm are stronger than reasons in favour of benefiting others, even when harms and benefits are equalized in magnitude. Note that HBA compares the moral reasons for and against two different types of doings. The contrast is between making it the case that someone is harmed and making it the case that someone is benefited. HBA is compatible with the view that your reasons against allowing someone to be harmed aren't stronger than your reasons in favour of allowing them to be benefited.

For reasons that will become clear soon enough, it will also be important for our argument to say something about the comparative strength of our reasons in favour of allowing others to be benefited. Presumably, we usually have reason to allow others to be benefited. But how do these reasons stand in comparison to our reasons in favour of actively benefiting other people?

You could hold the view that there is greater reason in favour of doing good than in favour of allowing good things to befall other people. You might even suppose that this difference is exactly equal to the difference in strength that obtains between our reasons against doing and allowing harm. Let's call this view the *Inverse Doctrine of Doing and Allowing* (IDDA). Although it might sound superficially attractive, we believe IDDA is untenable. To see why, consider the following case:

Medicine:

You are able to save one person by giving her a vial of some drug. A different person can save five people using the same vial. You cannot save these people, because you are not permitted entry to their location. Tragically, there is only one vial of the drug available. You can either use the vial to save one life or allow the other person to use the drug to save five lives.

What should you do? We think it's clear you ought to let the five be saved. But note the following. In *Rescue*, your reasons against killing one are stronger than your reasons against letting five die. If the IDDA were true, it would then seem to follow that your reasons in favour of saving one by your own hand are stronger than your reasons in favour of letting five be saved. But that conclusion is obviously unacceptable in a case like *Medicine*. So the IDDA is false. In fact, we would go so far as to say that your reasons in favour of benefiting

others are not any stronger than your reasons in favour of allowing people to be benefited, all else being equal.

Having clarified the non-consequentialist asymmetries that interest us, we next need to think about how to extend these principles to contexts in which outcomes are uncertain. We can think of the reasons quantified over in DDA and HBA as *objective moral reasons*: moral reasons grounded in the way things actually are, as opposed to the agent's evidence and/or credal state. Because the Paralysis Argument is addressed to agents who don't know what outcome they will bring about, we need some way of translating between reasons of this kind and *subjective moral reasons*, which we take to be grounded in the agent's evidence and/or credal state.

We follow the decision-theoretic approach to deriving subjective permissibility constraints from objective deontological principles proposed by Lazar (2018), drawing on past efforts to 'consequentialize' non-consequentialist moral theories (Broome 1991; Dreier 1993; Portmore 2011). On Lazar's approach, the decision-maker ranks the possible outcomes that she can realize in terms of the extent to which they would be favoured by her objective moral reasons, also known as their *choice-worthiness*. The outcomes are specified in such a way as to include everything that matters, morally speaking, be it the breaking of a promise, the intending of a harm, or the using of a victim as a mere means. The ranking should therefore not be confused with an ordering of outcomes in terms of their agent-neutral moral value. For example, the ranking will order the outcome in which the agent kills one to save five in *Rescue* below that in which she allows the five to die in order to avoid killing the one. We assume that the ranking is interval-scale measurable. Where the agent is uncertain about the outcome of her action, we then say that she has greatest subjective reason to perform that act which is best in expectation: i.e., the action that is most favourable when taking a

weighted average of the choice-worthiness of its possible outcomes, where each possible outcome is weighted by its subjective probability. Like Lazar (2018), we see no reason to think that the decision theoretic approach just described is inimical to the spirit of popular non-consequentialist moral theories.

Imagine, then, that there is some action, *D*, available to you, and some morally significant outcome, *O*, that could arise unpredictably as a result of your doing *D*, such that if you do *D* and *O* occurs and is harmful to some individual, then you will count as having done harm to that person by virtue of making it the case that *O* obtains. Consider, also, some alternative option, *A*, such that if *O* occurs subsequent to your choosing *A*, you will count as having merely allowed any harm involved in *O*. *D* might involve driving to the supermarket, whereas *A* might involve sitting motionless at home. The harm involved in *O* might be the unpredictable premature death of some person many years from now.

Suppose, now, that you have no greater reason to expect that the relevant harm will occur unpredictably as a result of choosing *D* or *A*. This seems very plausible, given that *O* is some unforeseeable event and you have no idea whatsoever about the causal route by which your behaviour might connect up with such an outcome. (Exactly how would your going out or staying home make the difference between someone many years from now dying prematurely or living a full life?) It would then seem to follow that you 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*.

In line with our discussion of IDDA, we think that if *O* obtains and is instead beneficial, then you do not have greater objective reason in favour of having brought about

that benefit as opposed to merely allowing it to occur. Certainly, if there is a discrepancy in the strength of these reasons, it is not nearly as great as that which obtains between doing and allowing harm. Just as it seems that you have no greater reason to expect that some harm will arise unpredictably as a result of choosing *D* or *A*, it would seem that you have no greater reason to expect some benefit to occur unforeseeably as a result of choosing *D* or *A*. Therefore, considering the possibility that *O* is beneficial fails to provide you with a reason for preferring *D* over *A* as strong as the subjective reason for preferring *A* over *D* arising in light of the possibility that *O* is harmful.

By iteration of this line of reasoning, we can derive the striking result that taking account of the indirect and unpredictable effects of your action gives you greater subjective reason to ensure, so far as is possible, that the indirect consequences of your behaviour are things you allow to happen, and not things you make happen. In other words, you have greater *pro tanto* reason to choose actions like *A* over actions like *D*.

What if we also take account of *D*'s direct and foreseeable consequences? If you know that *D* will foreseeably benefit someone significantly whereas *A* has no foreseeable benefits, might this not provide a reason in favour of choosing *D* sufficient to outweigh these reasons against?

This seems implausible to us in light of what we've already established about the significance of the indirect and unforeseeable effects of your actions. The long-run moral importance of anything you do appears to be determined primarily by effects of this kind. As we've noted, far more people are going to be harmed or benefited in ways you can't possibly foresee than are helped or hindered expectedly as a result of what you do. We think this is especially true of actions whose directly foreseeable effects are of great moral significance, such as saving a child's life. If you save a child's life, your life-saving action will be

indirectly responsible for every subsequent event in that life. It will be responsible for the existence of her children, and for everything they do, including the bearing of grandchildren. Each of these people will also impact on the reproductive choices of various other people outside their lineage, and the different people who come to exist as a result will subsequently change the composition of the next generation, and so on. The long-run future will be importantly different as a result of what you do, in ways you can't even begin to foresee. Given how long we should expect civilisation to last,[3] the overall moral significance of your action seems to hinge largely on indirect effects of this kind. Because of this, we think the balance of reasons is tilted decisively in favour of the indirect and the unpredictable.

The conclusion to which we're driven, then, is that you should try to do as little as possible. More precisely, you should try to ensure, so far as is possible, that the consequences of your behaviour are things you allow to happen, and not things you make happen. Instances of allowing can, in principle, involve lots of activity (Bennett 1995: 86-7, 137-8; Woollard 2015: 31-2). For example, most of us allow many children in the developing world to die of easily preventable causes while nonetheless keeping busy in our daily lives. Even so, it's hard to imagine that you could be especially active while also trying to ensure that as many consequences of your behaviour as possible are things you merely allow to happen.

A possible exception might arise if you were able to seal yourself off from the rest of the world in some fashion. Sealed away, you could then be as busy as you like, since nothing you do would disturb the outside world. But sealing yourself off in this way in the first place

[3] The Earth will remain hospitable to complex life for up to approximately 0.9 - 1.5 billion years, at which point the increasingly brighter ageing Sun will drive a catastrophic runaway greenhouse effect. However, if the human species spreads throughout the Universe, we will have approximately 100 trillion years before the stars die out (Adams 2008).

- buying a bomb shelter somewhere in Utah, travelling there, stacking it with provisions - seems to require acting in the world in just the way the argument apparently forbids. If so, there is no escape from paralysis.

3.

The previous section gave a detailed presentation of the Paralysis Argument. In this section, we'll focus on objections to the argument. Recall that we treat the Paralysis Argument as a *reductio*. The argument assumes HBA and DDA. If the truth of these principles would make the argument sound, we think this provides excellent grounds for rejecting their conjunction. The key question we'll examine, therefore, is whether we can find some other way to resist the Paralysis Argument.

3.1

The first objection we'll consider draws on Lenman (2000). According to Lenman, those effects of an action that are unforeseeable even to an ideally conscientious human agent are, as a rule, irrelevant to its status as permissible or impermissible. Of such consequences, Lenman says, "the agent should ordinarily simply not regard them as of moral concern." (363)

A view of this kind can easily seem intuitively compelling. How could effects of your behaviour beyond your ken impact on how we assess your conduct as a moral agent? Lenman suggests that disregarding such effects seems especially sensible when viewed in light of popular non-consequentialist views. According to these theories, "we should be morally engaged not by the quite futile project of promoting good long-term results but by more local

projects and concerns whereby, recognizing the fact of our epistemic limitations, we seek nonetheless to live virtuously, with dignity and mutual respect." (Lenman 2000: 364)

A key assumption of the Paralysis Argument is that taking account of the possible indirect and unforeseeable effects of some action can provide you with a reason against performing that action. Lenman may object that this assumption is one that non-consequentialists should anyway reject, and so the argument fails. However, we think Lenman's treatment of unforeseeable consequences is unconvincing. It seems impossible to draw a sensible distinction between the foreseeable and unforeseeable in light of which such a distinction could bear the moral significance placed on it by Lenman.

What does it mean for some outcome to be foreseeable (to an ideally conscientious human agent)? We can't say that some outcome of your action is foreseeable just in case you were (in principle) in a position to know that the outcome would follow from your choice of that act. There is certainly no plausibility in the suggestion that the potential for some action to bring about some consequence can provide a reason against performance of that action only if the agent is (in principle) in a position to know that the consequence will surely obtain if the action is performed. Any such view would permit reckless endangerment.

A more plausible view would suggest that some potential outcome of your action is foreseeable (in the morally relevant sense) if it is sufficiently probable conditional on performance of that action. But what could be meant by 'sufficiently probable'? Here is one thing we think this could not plausibly mean: a one-size-fits-all threshold falling somewhere on the probability scale such that *any* potential consequence can be morally disregarded just so long as its conditional probability is below the threshold.

To see why this is so implausible, consider

Button:

You have good reason to believe that pressing a certain button has the potential to kill one person, but will provide you with some modest benefit with certainty. Let p be the probability threshold below which potential consequences can be disregarded, and let the probability that pressing the button kills one person be $p - \varepsilon$, with ε an arbitrarily small positive quantity. You are just about to press the button, when you learn that in fact pressing the button has the potential to kill one million people with probability $p - \varepsilon$.

We think it's clear that you need to think again. Even if there is some threshold below which you can disregard the risk of killing one person, it seems absurd to suppose that the same threshold should be used for actions that have the potential to kill one million people. If Lenman were to propose that such a one-size-fits-all threshold should be used to delimit the supposedly morally significant boundary between the foreseeable and the unforeseeable, his view should be rejected.

The foregoing example suggests a better way forward. Whether some possible undesirable outcome is sufficiently probable that the agent can be required to forego an action because she might thereby bring about that outcome must depend not only on the probability of the outcome, but also the extent to which the outcome would be disfavoured by the agent's objective moral reasons (Kagan 1989: 87-91; Lazar 2018). But this is clearly fully compatible with the decision-theoretic approach to deriving subjective permissibility constraints from objective deontological principles that we relied on in setting out the Paralysis Argument. So there can be no objection to the argument on that basis.

3.2

Here is the second objection we'll consider.[4] In setting out the empirical foundations of the Paralysis Argument, we placed significant weight on the ubiquity of identity-affecting actions. We noted that almost anything we do can be expected to affect the timing of some reproductive event, leading to the existence of some person who would not otherwise have been conceived. Furthermore, we noted that this action will be causally responsible for everything this person does, for good or ill, and anything that is done by people who only exist as a result of this person's actions, and so on. However, we might think that when the causal sequence flowing from your behaviour to a harmful outcome passes through the voluntary actions of another agent, this counts against describing you as 'doing harm'. Thus, Woollard (2012) considers the idea that "bringing about harm through the voluntary action of another agent … might not count as harming." (685) Similarly, Quinn (1989) suggests that the DDA should be qualified so as not to apply to harms that arise from our activity if these are "remote from what we do" and "more directly traceable to the wrongful agency of persons more immediately concerned." (302) What should we make of this idea?

We grant that when one agent brings about harm to someone through the intervening agency of another, it will often be inappropriate to describe the first person as having done the harm. Foot (1984) notes that "having someone killed is not strictly *killing* him" (79). Thus, saying that Charles Manson killed many people can easily create the false impression that Manson himself wielded the knife in the Tate-LaBianca killings. More generally, locutions associated with the 'doing' side of the doing/allowing distinction may implicate direct involvement in harm. But this implication is contextually cancellable. Thus, when we say that Hitler and Stalin killed millions we take it as common knowledge that they acted through various intermediaries.

[4] We are grateful to an anonymous referee for raising this objection.

Of course, what matters to us is not how to use words, and Foot immediately adds that having someone killed "seems just the same morally speaking" (79). There may be no terms in ordinary English whose denotations and connotations map perfectly onto the distinction whose significance moral philosophers debate under the heading of 'doing/allowing.' If so, we should permit ourselves to warp the English language to suit the purposes of philosophical analysis (compare Bennett 1995: 3-7, 65-68). The question for us is whether harm that arises from our behaviour should sometimes be rejected from the 'doing' side of the doing/allowing distinction, so understood, when the causal chain runs through the voluntary agency of another person.

Once we understand the question in these terms, we think there is no genuine temptation to answer in the affirmative. It is clearly possible to do harm, in the relevant sense, even when that harm is more immediately traceable to some other person. Consider the following case:

Arms Trader

You are approached with the opportunity to sell a large volume of weaponry to a brutal dictatorship, foreseeing that the weapons will be used to oppress and murder innocent civilians. You also know that if you do not make the sale, the dictator will just go to one of your less scrupulous competitors and purchase the arms they want from them.

Intuitively, you should refuse to do business with the dictator in *Arms Trader*. This seems straightforward to explain in terms of the DDA. Although the dictator will purchase arms

elsewhere, you will not have been an active party to the suffering of her victims in that case. But this assumes that you can be actively involved in harming others in the way proscribed by the DDA, even when that harm is more directly traceable to the voluntary behaviour of some intermediary agent or agents.

We grant that it sometimes sounds wrong to say that you do harm to another when you initiate a causal sequence that ends with that person being harmed through the voluntary behaviour of some other agent. But so far as we can see, this is entirely explained in terms of pragmatic factors like those discussed earlier: i.e., in terms of conversational implicatures that typically attach to locutions associated with the 'doing' side of the doing/allowing distinction. Our reluctance to speak of 'doing harm' in these cases therefore provides no guidance as to what matters, morally speaking.

3.3

Other accounts of the doing/allowing distinction may impose special requirements on the sequence of intermediary events linking the agent's behaviour to the harm that occurs. Consider the theory developed by Woollard (2015). On her view, an agent counts as doing harm only if some fact about the agent's behaviour is part of a sequence leading to the harm. In order for a fact to form part of a sequence leading to a harm, it must be *substantial* or *relatively substantial*, where a fact's being *positive* (roughly, telling us what *is* the case, as opposed to what *is not*)[5] is a paradigmatic way of being substantial, whereas a fact describing the absence of a barrier to harm belonging to the victim is a paradigmatic way of being relatively substantial. To count as doing harm, every link in the chain of facts connecting the

[5] See section 3.6 for more on the nature of positive facts.

agent's behaviour to the harm must be substantial or relatively substantial, not just the fact that describes the agent's behaviour (Woollard 2015: 30).

When it comes to present acts that result in unforeseeable future harms, we may wonder whether we can count on every single fact linking current behaviour to future harm satisfying this criterion. How likely is it that the distant, unpredictable effects of our actions will be linked to us via unbroken chains of positive facts, depending not at all on facts about something not happening or someone or something not being present somewhere?[6]

In the first instance, we should note that being positive is merely one way in which a fact can form part of a sequence that links the agent's behaviour and the victim's harm such that the agent counts as doing harm to victim. On Woollard's account, there are a variety of ways in which negative facts can count as part of sequences associated with harms that are done and not merely allowed (see Woollard 2015: 21-79). Her account would otherwise be very implausible. We do not want to say that a person who kills someone else by stopping her heart from beating does not count as doing harm in killing the victim just because the absence of a heartbeat is a negative fact.

Nonetheless, given that the causal chains linking present acts and unforeseeable future harms are so convoluted, isn't it unlikely that every link will be substantial and/or relatively substantial, even according to a more expansive conception that allows negative facts like the absence of a heartbeat to count as part of a sequence? Because the links in the chain are inscrutable, for all we know, any one of them could involve a fact that fails to qualify as substantial or as relatively substantial. Because each link has some probability of failing to qualify as such and there are so very many links to be considered, surely the probability that

[6] We are grateful to an anonymous referee for raising this concern.

some link or other is insubstantial is very high. Therefore, we may think, it is very unlikely that we can be counted as doing harm in this way, according to Woollard's theory.

Note that this line of reasoning does not actually depend on the details of Woollard's theory. It can be run given any theory of the doing/allowing distinction that imposes special requirements on the sequence of intermediary events linking the agent's behaviour to the harm that occurs if the agent is to count as doing harm. Given that the causal chains linking present acts and unforeseeable future harms are so convoluted, for all we know, any link in the chain could fail to satisfy whatever requirement is imposed by the theory. Because each link has some probability of failing to satisfy the requirement and there are so very many links, we might similarly conclude that the probability that at least one link fails to do so is very high.

In response to this concern, we invite the reader to consider the following thought experiment:

Mystery Box

The illusionist Fantastico presents his audience with an enormous, opaque box and invites you up on stage. Inside the box is a mysterious, convoluted machine. Fantastico lets you know that the machine is extremely complicated, involving the operation of millions of distinct mechanisms. But that is all he knows. Exactly what goes on inside the box is a complete mystery. Even so, he lets you know that if you press a button located on top of the box, the machinery will kick into gear, and once it has concluded its convoluted operation, something will happen and the person seated in the middle of the first row will be left dead.

Clearly, if you press the button and the person seated in the middle of the first row ends up dead, then you will count as doing harm by killing her. But this is a case in which the causal

chain linking the pressing of the button to the harm of the victim is convoluted and inscrutable. We have no idea what goes on inside the box.

There must be something wrong with trying to derive the conclusion that harm that results from a convoluted and inscrutable causal chain cannot with high confidence be counted as harm that the person initiating the chain makes happen due to the theoretical imposition of a requirement that each link in the chain must satisfy but that for all we know is not satisfied by any particular link. We do not wish to take a firm stance on exactly where the derivation goes wrong: on whether the problem arises from inferring a high probability that some link or other fails to satisfy whatever requirement on intermediary events is imposed or from imposing such a requirement in the first place. But something must go wrong somewhere.

3.4

Here is a fourth objection. It may be suggested that when your action unforeseeably harms someone further down the line, this act is nonetheless justifiable if it was *ex ante* weakly Pareto optimal: that is, if the action was, in expectation, at least as good as every other alternative for every individual affected, and strictly better than every other alternative for at least one affected individual.

Consider, again, the action of driving to the supermarket. Suppose you expect that this action will have significant unforeseeable knock-on effects for the reasons we noted in section 2. Even so, looking to its indirect effects, this action was presumably no more likely to harm than to benefit whomever it might in fact end up harming. After all, the causal sequences involved here are completely inscrutable. For any particular individual living now or in future, it's just as easy to imagine convoluted chains of events by which this person ends

up being benefited unforeseeably by your action as ones by which they end up being harmed. There doesn't seem to be anyone who is expectedly worse off as a result of your action — and you are presumably better off. Therefore, it seems plausible that driving to the supermarket was *ex ante* Pareto optimal. Because the action was *ex ante* Pareto optimal, we might reason that it must have been justifiable to each person: from the epistemic position obtaining at the point at which you chose to act, no one could reasonably reject a principle permitting this action, since everyone was at least as well off in expectation (compare Frick 2015a: 186-191). Being justifiable to each person, your action was therefore permissible. We might go so far as to say that the status of some action as *ex ante* weakly Pareto optimal generally overrules or undermines the force of familiar deontological constraints on doing harm, by virtue of implying that the action is justifiable to all on grounds they could not reasonably reject from an *ex ante* perspective.

Of course, the act of driving to the supermarket was one we selected arbitrarily. If the reasoning set out above can be used to justify this action even when we take account of its potential to bring about indirect and unforeseeable harms, then similar reasoning can presumably be used in general to justify going about our day-to-day lives in the way we're used to.

However, we think that non-consequentialists ought to reject the idea that *ex ante* Pareto optimality vitiates the force of deontological reasons against doing harm. Consider this variation on *Rescue*, due to Kamm (1996: 303):

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.

It might be the case that because the ambulance will run over a pedestrian only when necessary to save the greater number and all people in the community are equally likely to walk on its roads and to need medical help, running the community's ambulance service using this machine is *ex ante* Pareto optimal. Nonetheless, we take it that the intuitive thing to say about this case is that the community may not run an ambulance service using this sort of vehicle. Moreover, the reason for this seems, intuitively, to be the existence of the same kind of constraint on doing harm in play in *Rescue*. In this case, however, the harm that expectedly eventuates need not be the result of any morally responsible agent making a decision that was not *ex ante* Pareto optimal at the time at which it was made, because the harm that will be done (if it is done) will be under the control of the on-board AI.

You might reply as follows. Since it would be impermissible for a human driver of the ambulance to run over one person in order to save five, it seems natural to imagine that the community in *Ambulance* would consider the purchase of the ambulance only insofar as they think this will allow them to get away with this sort of thing, by outsourcing the decision to a machine who lacks moral responsibility. This may be thought to play a key role in explaining why it would be wrong to run a fleet of ambulances of this kind. The on-board AI is intended as a surrogate for a human agent, and we may insist that any surrogate of this kind must be bound by similar moral constraints as bind the human agents they replace (compare Frick 2015b: 210-11).

We don't find this reply convincing. The objectionable character of making the purchase in *Ambulance* doesn't hinge on interpolating details into the case in light of which

the on-board AI is intended as a surrogate for a human agent. For example, we need not conceive of the community as willing to purchase the special ambulance only because they think this will allow them to circumvent the deontological constraint on harming others. It might simply be more desirable for other reasons. Perhaps it ordinarily allows for more efficient transport of patients. We can also imagine that the community inhabits a world where all motor vehicles are self-driving by legal requirement. No on-board AI would therefore be considered a surrogate for a human being. Intuitively, none of this makes any difference.

3.5

Here is a fifth objection: that, perhaps, the Paralysis Argument is undermined by the Non-Identity Problem (Parfit 1984: 351-79). The Paralysis Argument leans crucially on the premise that the overall moral importance of anything you do is determined primarily by its indirect effects. Far more people are harmed or benefited indirectly in ways you can't possibly foresee than are helped or hindered expectedly as a result of what you do. The Non-Identity Problem may be thought to set a limit on just how many people can be indirectly harmed or benefited by your actions.

By the Counterfactual Comparative Account (CCA) of harm that drives the Non-Identity Problem, a person is harmed by some action only if they would have been better off had the act not been performed. Any person whom you harm must therefore exist both in the actual world and in the nearest possible world in which you fail to perform the act in question. Call any such person 'necessary' relative to the act in question (Österberg 1996). Given that the potential for even our most mundane actions to have far-reaching consequences depends on their capacity to affect who exists in the future, it may be objected

that ‘necessary’ people will quickly become scarce. In this way, the Non-Identity Problem may be thought to rein in the overwhelming moral significance that otherwise seems to attach to the long-run impact of your behaviour, capping the number of people who can be harmed or benefited unforeseeably. And this may be thought to shift the balance of moral reasons back toward the immediate and foreseeable.

We see two problems with this objection. Firstly, it will only appeal to philosophers who endorse CCA. Many reject it (Hanser 2008, 2009; Harman 2004, 2009; Shiffrin 1999; Woollard 2012). Our impression is that CCA appeals mostly to consequentialists (like Jackson 1997 and Norcross 2005), with deontological theorists tending toward greater scepticism, pushing back, in particular, on its application to the Non-Identity Problem. And anyone who endorses CCA will face challenges when it comes to explaining why it seems wrong to bring about overdetermined harm in cases like *Arms Trader*.

Secondly, we see no good reason to accept the suggestion that ‘necessary’ persons will become scarce quickly enough for the Paralysis Argument to come undone. The objection we’re discussing suggests that when we consider those ‘necessary’ people who will be harmed or benefited indirectly as a result of some action, the numbers drop off very quickly with time. That doesn’t seem plausible to us. The emphasis in the Paralysis Argument is not on harms or benefits that accrue to those people who exist unforeseeably as a result of the identity-affecting character of our actions, but rather the harms or benefits that others receive as a result of the fact that these people exist. These other people do not depend for their existence on our behaviour, at least not within any suitably short timeframe. We expect it would require hundreds or thousands of years before the morally significant consequences of some action of yours have become so far-reaching that everyone affected by that action also exist as a result of its performance.

3.6

Here is the final objection we'll discuss, which focuses on the question of what it means to be minimize the extent to which the morally significant effects of your behaviour are things you make happen.

We took the Paralysis Argument to support doing as little as you can: for example, sitting motionless. Sitting motionless, we assumed, ensures - to the greatest extent possible - that the consequences of your behaviour are things you allow to happen, and not things you make happen. This seems intuitive. However, failing to act does not guarantee that you will count as merely allowing some outcome to occur (Foot 1967: 11; Bennett 1995: 96-100, 112-4; Woollard 2015: 47 - 51).

Here is one case that illustrates this possibility, due to Bennett (1995: 98):

Earthquake

An earthquake knocks you over, so that you fall on top of someone else, putting pressure on her chest. If you remain lying across her chest, she will be unable to breathe and will suffocate.

Here it seems implausible to suppose that you would merely allow this person to die if you failed to move. Your passivity would certainly not have the moral quality of a mere allowing. Ordinarily, it's permissible to allow one to die to save five others, but if you knew in *Earthquake* that remaining still during this period would save five people from death, intuitively, you would still not be permitted to suffocate this person with the weight of your

body. Furthermore, in moving from this person's chest, we think it makes most sense to say that you would merely be allowing the five to die by omitting to do what was necessary for their survival.

Remaining motionless therefore does not guarantee that the consequences of your behaviour are merely things you allow to occur, as opposed to things you make happen. There are cases in which that aim is better achieved through activity as opposed to inactivity. As a result, we may wonder whether the Paralysis Argument really does entail paralysis. Given that a requirement to minimize doing harm can in principle obligate us to do anything *but* remain motionless, perhaps that requirement need not be especially onerous when applied to our daily lives in light of the Paralysis Argument.

Questions about the relationship between immobility and the doing/allowing distinction have played a prominent role in the discussion surrounding Jonathan Bennett's analysis of that distinction. Bennett (1995) initially suggested that you make something happen iff your behaviour is *positively relevant* to the outcome, and you allow something to happen iff your behaviour is *negatively relevant*. Your behaviour is positively relevant to some outcome iff the weakest fact about your behaviour needed to complete an explanation of the outcome is a *positive fact*. Iff the weakest fact about your behaviour needed to complete an explanation of the outcome is a *negative fact,* then your behaviour is negatively relevant instead. A fact about your behaviour is positive iff it rules out the (vast) majority of the options in your *behaviour space*: the space of the all the ways you could have moved. A fact about your behaviour is negative iff it fails to rule out the (vast) majority of options in your behaviour space. It follows from this that if the weakest fact about an agent's behaviour needed to explain some event is that the agent remained motionless, then the agent counts as making the event happen, as opposed to allowing it to happen. After all, remaining

motionless ordinarily rules out the vast majority of the options in the agent's behaviour space. Therefore, outcomes that depend specifically on you holding motionless are outcomes that you make happen. Typically, immobility is not the weakest fact about your behaviour needed to explain some outcome. A positive fact specifying that you performed some particular set of movements or a negative fact specifying that you *did not* perform some particular set of movements is more standard. But in some cases - like *Earthquake* - immobility is the weakest fact needed to explain the outcome. In those cases, immobility counts as making something happen.

This analysis may seem to have important implications for how we think about the Paralysis Argument. If you were to choose right now to just sit where you are for as long as you physically can, then some very particular future history will unfold that would not have developed similarly had you gone out into the world. On Bennett's analysis, your motionlessness will count as positively relevant to this long-run sequence of outcomes. Because what unfolds as a result of your holding still depends so minutely on the decision to hold still, we might infer that concern to minimize the extent to which the morally significant indirect effects of your behaviour constitute doings, as opposed to allowings, cannot plausibly recommend this course of inaction.

In fact, Bennett's analysis may seem to support the view that insofar as you are concerned to minimize the extent to which the morally significant effects of your behaviour constitute mere allowings, the indirect and unforeseeable effects of your actions should not concern you at all. We have emphasized the extent to which the constituents of the future population are extremely sensitive to apparently insignificant actions and inactions that we undertake in the here and now. Because of effects of this kind, when we look to the indirect effects of our behaviour, we find some very particular future history unfolding that would not

otherwise have occurred. On Bennett's analysis, that history is therefore a sequence of events with respect to which our behaviour is positively relevant, and therefore something we make happen, whatever we happen to be doing (or not doing). As a result, we may conclude that the morally significant indirect effects of our behaviour can provide no grounds for preferring one option over another insofar as you are trying to minimize the extent to which the morally significant indirect effects of your behaviour constitute doings, as opposed to allowings.

We have two objections to this line of argument. Firstly, even if the act of sitting still may be positively relevant to the occurrence of some particular future history, it need not follow that the act is positively relevant to the occurrence of any particular event within that history. Positive relevance does not distribute across conjunctions. Suppose that if you do *A*, then *e*1, *e*2, and *e*3 will occur. There are many other actions that you can perform: *B, C, D*, *etc*. None of these results in all three events: one third result in *e*1 and *e*2 but not *e*3, one third result in *e*2 and *e*3 but not *e*1, and one third result in *e*1 and *e*3 but not *e*2. Hence, performance of *A* has positive relevance with respect to neither *e*1, nor *e*2, nor *e*3, since each event would nonetheless have occurred under the majority of alternative options available to you. However, their conjunction depends specifically on performance of *A*, so *A* is positively relevant to their joint occurrence. Similarly, if some particular future population is realized only if you remain still, it need not follow that any individual constituent of that future population depends for their existence on your motionlessness. You might well count as allowing each of these people, considered individually, to come into existence, and as allowing every individual causal upshot of their existence, even if you make it the case that they as a group exist, and also that the sum total of the causal upshots of their existence come to bear. This idea is admittedly pretty hard to get your head around. At the very least, attention to this point shows that it is quite unclear that concern to minimize the extent to

which the morally significant effects of your behaviour constitute things you make happen should not favour motionlessness, even if we accept Bennett's analysis and grant that some particular future history would depend positively on your immobility.[7]

Our second objection is that Bennett's analysis is often attacked precisely because it handles immobility in the way we have described. Many people are highly sceptical of the idea that every case in which the weakest fact about an agent's behaviour needed to explain some event is that the agent remained motionless represents a case of doing, as opposed to allowing (Dinello 1994; Locke 1982; Quinn 1989). Consider this well-known example:

Explosion

If Henry remains motionless inside the sealed room he is in, some electric dust will fall in such a way as to close a tiny electric circuit, setting off an explosion that kills Bill.

Suppose Henry lies still, disinterestedly daydreaming as the dust settles. It doesn't sound right to say that he thereby killed Bill, as opposed to allowing Bill to die. Bennett's analysis suggests the opposite, since Henry's immobility makes his behaviour positively relevant to the event of Bill's death.[8]

Furthermore, consider the following modification of *Explosion*, due to Quinn (1989)

[7] We're grateful to [redacted] for this observation.

[8] In response, Bennett (1995: 99, 111-24) suggests that our intuitions about *Explosion* hinge on whether Henry must exert himself in order to remain still. If Henry sweats and strains to keep absolutely motionless because he wants Bill to die, it sounds more intuitive to suppose that Henry didn't merely allow Bill to perish in the explosion. Remaining motionless in response to hearing the Paralysis Argument would presumably also require significant effort, so the addition of this epicycle to Bennett's analysis would not defuse any challenge posed to the Paralysis Argument.

*Explosion**

If Henry remains motionless inside the sealed room he is in, two things will happen: (a) some electric dust will fall in such a way as to close a tiny electric circuit, setting off an explosion that kills Bill; (b) five people who are falling from a very great height will be saved by being caught in the net that Henry is holding, which must be kept in this exact position.

Like Quinn, we think Henry is permitted to remain motionless in this case, however much he might need to sweat and strain in order to achieve this. Typically, killing one to save five is impermissible, as in *Rescue*. This implies that remaining motionless in this case *either* does not constitute a case of doing harm after all *or* else represents a case of doing harm that is sufficiently different from paradigm instances of killing and sufficiently similar to paradigm instances of letting die that it takes on the moral quality of an allowing, as Kamm (1996: 27-8) suggests. Regardless of which interpretation we choose, we end up with the conclusion that immobility of this kind is immobility of a kind that the Paralysis Argument may recommend in light of the possible indirect harms we could bring about through acting in the world.

We remain unsure about exactly what feature of *Earthquake* makes this a case of killing through inaction. Because of this, we're not totally confident that the Paralysis Argument supports immobility. It's our sense, however, that the onus is on non-consequentialists to argue the case. The default expectation is surely that inactivity counts as allowing. The existence of cases like *Earthquake* doesn't seem to change that fact. It's natural to think that *Earthquake* is somehow unusual.

We also note the following. If the Paralysis Argument doesn't require you to sit motionless, it probably requires something else of you that is only slightly less undesirable. There may be some other deeply unappealing option that ensures, so far as is possible, that the indirect effects of your behaviour are things you allow to happen, and not things you make happen. Perhaps you have to live in such a way as to exercise as little agency as possible, assenting to every impulse, never planning, never intending. Being constrained to live in this way seems only slightly more inviting than sitting motionless. Therefore, the Paralysis Argument would still function as a *reductio*, even if it doesn't require you to just sit still and let the world pass you by.

3.7

Before we present our favoured response to the Paralysis Argument, we round out this section with the following methodological observation, which seems especially pertinent in light of our discussion in section 3.6.[9]

The nature of the doing/allowing distinction is obviously contested. In constructing the Paralysis Argument, we have tried to stick as close as possible to intuitive, pre-theoretic judgments about the nature and application of the distinction, so as to avoid presupposing any particular philosophical theory on the matter. However, this pre-theoretic understanding is undoubtedly problematic in many ways. It is the job of a philosophical analysis to straighten out the problematic features of our ordinary understanding, in a way that may require revision of certain pre-theoretic judgments that turn out to be unsupportable in reflective equilibrium. The pre-theoretic beliefs to which philosophical analysis addresses itself are always laden

[9] We're grateful to an anonymous referee for inviting us to discuss these points.

with philosophically weighty presuppositions that some philosophers believe should have no part in a more rational theoretical construction.

In this respect, our handling of the doing/allowing distinction arguably presupposes a great many things that are philosophically controversial. We have touched on a few of these presuppositions in this section, and have tried to defend them. There are probably others that have escaped our notice. We think the burden of proof generally falls on the philosopher who challenges our pre-theoretic commitments, and so we think the burden of proof falls on the proponents of particular theories to refute the Paralysis Argument by showing it to be based on a flawed understanding of the doing/allowing distinction. However, we personally think the best way of responding to the argument takes a very different form, which we now proceed to outline.

4.

In this section, we'll consider one last way in which someone might respond to the Paralysis Argument. This response turns on noting that the argument seems to have assumed that our actions are not already oriented toward producing good effects over the very long run.

In setting out the argument, we contrasted the moral significance of the direct and foreseeable consequences of your behaviour with those of its indirect and unforeseeable effects. Even if the direct and foreseeable consequences of your actions are morally very important, they probably aren't as important as the indirect and unforeseeable effects, given how long civilisation will last. So, even if the direct effects of an action confer great benefits, you still shouldn't act because the long-run harms are so much greater. But this line of reasoning seems to assume that insofar as your actions aim at some morally important

outcome, that outcome is near-term. What if your actions instead aim at improving the whole of the very long-term future?

For example, what if you are trying to reduce the risk of human extinction by working to mitigate the risk of a catastrophic bioengineered pandemic, or improving the long-run condition of humanity by working to mitigate climate change? Then what there is to be said in favour of your action may be thought to be some potential effect on the long-run development of civilization, and not merely some foreseeable near-term benefit. In that case, we can't argue as above. We are no longer weighing the short-term against the long-term. Instead, everything comes down to the probable and not so probable effects of your behaviour on the long-run future. Therefore, it seems, the Paralysis Argument fails to support a prohibition on acting in the world when our behaviour has this kind of long-term focus.

Even so, the line of reasoning set out above at best yields permissions to engage in long-term oriented actions. It provides no defence of actions that aren't geared toward improving the long-run future of humanity. Unless your entire life is oriented toward this goal, you won't be permitted to do much of anything. To escape paralysis, your every motion must be at the service of posterity. This seems extraordinarily demanding.

This need not be a decisive objection, at least not insofar as we are comparing the ability of consequentialist and non-consequentialist theories to adequately capture the moral significance of the indirect and unforeseeable effects of our actions. Suppose it's true that a non-consequentialist hoping to resist the Paralysis Argument will be driven to the view that our lives should be dominated by efforts to reduce the risk of human extinction or otherwise positively shape the long-term future. Well, consequentialism *also* entails that all your life should be oriented toward ensuring a good long-run future for humanity. The argument for this is straightforward given what we know about the potential size of the future, so long as

we assume a population axiology on which the addition of flourishing lives makes for an intrinsically better outcome (on which see Broome 2005, Huemer 2008). Because there could potentially exist so very many flourishing lives in the future, actions that reduce the risk of human extinction or otherwise positively shape the long term future are assigned enormous expected value (Beckstead 2013; Bostrom 2003). Based on conservative estimates about the possible size of the future population, Bostrom (2013) calculates that "the expected value of reducing existential risk by a mere *one millionth of one percentage point* is at least a hundred times the value of a million human lives." (18-19) What we have, then, is a striking convergence in respect of how consequentialists and non-consequentialists should take account of the moral significance of the long-run future. It is as if they have been climbing the same mountain, but from different sides (compare Parfit 2011: 418-9).

Perhaps very few non-consequentialists will see the recommendation to live a long-termist life as an inviting response to the Paralysis Argument. Taking this way out will obviously undercut the demandingness objection to consequentialism once and for all. We're not too concerned about this. We have never seen very much merit in that objection. Insofar as we are sceptical of consequentialism, we are more concerned about its failure to recognize constraints. If the argument of this paper is correct, non-consequentialists who endorse constraints may be forced to give up the demandingness objection as a consideration favouring their view.

5.

Standard non-consequentialist theories endorses asymmetries between harming versus benefitting and doing versus allowing. These asymmetries, combined with empirical facts about our actions' indirect impact on the long-run future, lead to the conclusion that the only

permissible courses of action either involve doing as little as possible or dedicating one's life to improving the very long-run future of civilisation. The non-consequentialist can either accept this implication, at the cost of making morality extremely demanding, or revise the usual account of the asymmetries between harming versus benefitting and doing versus allowing. Either way, some core aspect of the standard non-consequentialist understanding of morality has got to go.

Bibliography

Adams, Fred C. (2008) Long-term astrophysical processes. In Bostrom and Cirkovic, eds. *Global catastrophic risks*, 33-47. Oxford: Oxford University Press.

Beckstead, Nick (2013) *On the overwhelming importance of shaping the far future*. PhD thesis: Department of Philosophy, Rutgers University. <https://rucore.libraries.rutgers.edu/rutgers-lib/40469/PDF/1/play/>

Bennett, Jonathan (1995) *The act itself*. Oxford: Oxford University Press.

Bostrom, Nick (2003) Astronomical waste: the opportunity cost of delayed technological development. *Utilitas* 15, 308-14.

Broome (1991) *Weighing goods: uncertainty, equality and time*. Oxford: Blackwell.

(2005) Should we value population? *Journal of Political Philosophy* 13, 399-413.

Dinello, Daniel (1994) On killing and letting die. *Analysis* 31, 84-6.

Dreier, Jamie (1993) Structures of normative theories. *The Monist* 76, 22-40.

Foot, Philippa (1967) The problem of abortion and the Doctrine of Double Effect. *Oxford Review* 5, 5-15.

(1984) Killing and letting die. In Garfield and Hennessy, eds. *Abortion: moral and legal perspectives*, 177-185. Amherst, MA: the University of Massachusetts Press.

Frick, Johann (2015a) Treatment versus prevention in the fight against HIV/AIDS and the problem of identified versus statistical lives. In Cohen, Daniels, and Eyal eds. *Identified versus statistical lives: an interdisciplinary perspective*, 182-202. Oxford: Oxford University Press.

(2015b) Contractualism and social risk. *Philosophy and Public Affairs* 43, 175-223.

Gleick, James (1987) *Chaos: making a new science*. London: Vintage.

Greaves, Hilary (2016) Cluelessness. *Proceedings of the Aristotelian Society* 116, 311-339

Hanser, Matthew (2008) The metaphysics of harm. *Philosophy and Phenomenological Research* 77, 421-50.

(2009) Harming and procreating. In Wasserman and Roberts, eds. *Harming future persons*, 179-99. Dordrecht: Springer.

Harman, Elizabeth (2004) Can we harm and benefit in creating? *Philosophical Perspectives* 18, 89-113.

(2009) Harming as causing harm. In Wasserman and Roberts, eds. *Harming future persons*, 137-154. Dordrecht: Springer.

Huemer, Michael (2008) In defence of repugnance. *Mind* 117, 899-933.

Jackson, Frank (1997) Which effects? In Dancy, ed. *Reading Parfit*, 42-53. Oxford: Blackwell.

Kagan, Shelly (1989) *The limits of morality*. Oxford: Oxford University Press.

Kamm, F. M. (1996) *Morality, mortality, vol. 2: rights, duties, and status*. Oxford: Oxford University Press.

(2015) *The trolley problem mysteries*. Princeton, NJ: Princeton University Press.

Lazar, Seth (2018) In dubious battle: uncertainty and the ethics of killing. *Philosophical Studies* 175, 859-83.

Lenman, James (2000) Consequentialism and cluelessness. *Philosophy and Public Affairs* 29, 342-70.

Locke, Don (1982) The choice between lives. *Philosophy* 57, 453-75.

McMahan, Jeff (1993) Killing, letting die, and withdrawing aid. *Ethics* 103, 250-79.

Norcross, Alastair (2005) Harming in context. *Philosophical Studies* 123, 149-73.

Nye, Howard (2014) Chaos and constraints. In Boersema, ed. *Dimensions of moral agency*, 14-29. Cambridge: Cambridge University Press.

Österberg, Jan (1996) Value and existence: the problem of future generations. In Lindström, Sliwinski, and Österberg, eds. *Odds and ends*, 94-107. Uppsala: Uppsala Universitet.

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

Portmore, Douglas (2011) *Commonsense consequentialism*. Oxford: Oxford University Press.

Quinn, Warren (1989) Actions, intentions, and consequences: the Doctrine of Doing and Allowing. *Philosophical Review* 98, 287-312.

Scheffler, Samuel (1982) *The rejection of consequentialism: a philosophical investigation of the considerations underlying rival moral conceptions.* Oxford: Oxford University Press.

Shiffrin, Seana (1999) Wrongful life, procreative responsibility, and the significance of harm. *Legal Theory* 5, 117-48.

Singer, P. (1972). Famine, affluence, and morality. *Philosophy & public affairs*, 229-243.

Smith, Peter (1998) *Explaining chaos*. Cambridge: Cambridge University Press.

Thomson, Judith Jarvis (1985) The trolley problem. *The Yale Law Journal* 94, 1395-1415.

Toups, Des (2011) How many times will you crash your car? *Forbes* <https://www.forbes.com/sites/moneybuilder/2011/07/27/how-many-times-will-you-crash-your-car/#399122754e62> Accessed 30/04/2019/

Webb, Mark (2018) Jain philosophy. *The Internet Encyclopedia of Philosophy*, ISSN 2161-0002, <http://www.iep.utm.edu/jain/> Accessed: 11/07/2018.

Woollard, Fiona (2012) Have we solved the Non-Identity Problem? *Ethical Theory and Moral Practice* 15, 677-90.

(2015) *Doing and allowing harm*. Oxford: Oxford University Press.

Woollard, Fiona and Howard-Snyder, Frances (2016) Doing vs. allowing harm. In Zalta, ed. *The Stanford Encyclopedia of Philosophy*, <https://plato.stanford.edu/archives/win2016/entries/doing-allowing/> Accessed 11/07/2018.