DOI: 10.1111/phpr.70053

Philosophy and Phenomenological Research

ORIGINAL ARTICLE

Killing, letting die, and normative inertia

Tim Henning

Institute of Philosophy, University of Mainz, Mainz, Germany

Correspondence
Tim Henning, Johannes Gutenberg-Universität, Philosophisches Seminar, D-55099 Mainz, Germany.
Email: thenning@uni-mainz.de

Abstract

According to commonsense and deontological ethics, it is impermissible to kill a person in order to save the life of another person, all else being equal. But why? This article suggests a justification of this restriction. It appeals to the idea of *normative inertia*—very roughly, to the idea that in practical decisions, changing course is harder to justify than staying on track. So, in a nutshell, the reason why we may not kill one person to save another is not that killing is worse than letting die, but that it is *not better*. Various arguments for, and various formulations of, principles of normative inertia are discussed. The article then argues for a specific principle that is not directly concerned with moral permissibility but with practical rationality in a general sense. When coupled with an independently motivated moral premise, this principle can be shown to vindicate intuitions about restrictions on killing. The resulting account is not only plausible and well-motivated. Further, it vindicates intuitions about cases of "withdrawing aid" that have long seemed puzzling. Finally, the account presented here fits particularly well into a contractualist moral theory, refuting the widely held view that contractualism cannot accommodate deontic restrictions.

According to commonsense and deontological ethics, there are deontic restrictions on doing harm.[1] One such restriction forbids the killing of a person even if doing so would save another person, and all else is equal. This article presents a moral justification for the view that this restriction holds.

The proposed justification appeals to the idea of *normative inertia*—very roughly, to the idea that changing course is generally harder to justify than staying on track. Malm (1989) and Sartorio (2008) propose that morality is governed by normative inertia, and show that this leads to a new perspective on the restriction on killing. Unfortunately, they do not give a convincing explanation of *why* there should be inertia. Nor do they offer a precise formulation of the idea. This article aims to fill these lacunae. It argues that the relevant kind of inertia is only indirectly a moral matter. Instead, the inertia principle in question is a general principle of practical rationality. I offer a precise formulation, and I explain how it lends itself to a defense of a moral restriction on killing.

One important feature of the resulting account: It does not entail that letting die is *generally* less problematic than killing. For example, the account does not support the idea that our failure to save poor people in other countries from death through illness or starvation is less bad than killing them. So my defense of this deontic restriction cannot serve as a bulwark against the demanding implications of certain ethical views. I leave it to readers to decide if this is a flaw in my account.

A second feature of my view is worth noting in advance. In the course of my argument, I will show that my proposal can be naturally embedded in a contractualist moral framework. This may be welcome news for (some) contractualists. After all, contractualism is often claimed to be incapable of accommodating deontic restrictions.[2] I show that contractualism has more resources.

Section 1 starts with examples to highlight the intuitive moral difference between killing and letting die, and to illustrate the philosophical issues raised by the relevant intuitive judgments. Section 2 introduces the idea of normative inertia, highlighting the important contributions of Malm and Sartorio, but also explaining where they leave open questions. Section 3 turns to my defense of a normative inertia principle for rational decision-making. It is stated with some precision and given a pragmatic defense. The central idea is that rational agents avoid excessive deliberation costs, and that to do so they should follow the rule to make changes only if these changes are expected to be, not just equally good but *for the better*. Section 4 formulates a moral premise, according to which it is wrong to treat people as if their lives were of inferior value compared to lives of other persons, so that it is a better thing if they die rather than those others. This premise is independently plausible, but it is also given a contractualist defense. Section 5 then shows how these two ingredients, the principle of practical rationality and the moral premise, yield an argument for the restriction on killing that was illustrated in Section 1. Section 6 shows that my view solves notorious issues of withdrawing aid. Section 7 mentions some limitations and open questions and concludes.

1 | RESTRICTIONS ON KILLING

The restrictions that are in the focus of this essay are restrictions on killing. (However, the argument will be general enough to cover other cases of harming vs. allowing harm.) They prohibit killing even if killing would save an equal number of people from dying; in fact, they prohibit

[1] There is a wealth of virtually synonymous terms in the literature (constraints, side-constraints, etc.). Many of these would serve my purposes equally well.

[2] References are provided in Section 4 of this paper.

killing even if killing prevents an equal number of people from *being killed*—by another, or by the same person.

Throughout, I make heavy use of many moral philosophers' favorite tool: trolleys. Here goes:

> *Trolley K/LD* ("killing vs. letting die")
>
> An out-of-control trolley is running along a track, where it will eventually kill one person. You could use a switch to redirect it onto a sidetrack where it would kill one other person.

One remark: Unless I explicitly note otherwise, I ask you to assume that there are no differences between the parties that you consider relevant. If, for example, you think that differences in life expectancy, desert, or personal ties, etc., might make it the case that more is to be said in favor of sparing one party, I ask you to assume that no such differences are present. Another remark: Unless noted otherwise, my argument will be neutral on the question of whether, absent relevant differences, the two outcomes are determinately equally good, or whether they are incommensurable or "on a par."

Many philosophers believe that in this case, using the switch and deflecting the harm would be morally impermissible, and I share this judgment.[3] This case is often considered to illustrate a general constraint on doing versus allowing harm (e.g., McMahan, 1994). In fact, some think it is an ideal test case for the existence of such a constraint (Malm, 1989). After all, everything else is equalized: Neither of the deaths, we have assumed, would involve greater harm; redirecting would not involve an intention to kill; nor would any victim be used as a means; etc. Other, more subtle distinctions do not seem to play a role either.[4] The only salient difference is the mode of agency.

If the judgment is correct and redirecting in Trolley K/LD is morally impermissible, this is an instance of the restriction on killing characterized above. Moving on, here is a second example:

> *Trolley K/K*
>
> An out-of-control trolley is running along a track, where it will eventually kill one person. The trolley has been set in motion by a villain in order to kill this person. You could use a switch to redirect it onto a sidetrack where it would kill one other person.

Thomson (1991, p. 309), who presents this case, judges that you may not redirect the trolley. I agree: It seems wrong for you to redirect and kill an innocent person, even if this prevents someone else from also killing an innocent person. I suspect many other defenders of deontic restrictions will agree. As I said, many of them are on record saying that you may not redirect in the previous

[3] For example, McMahan (1994) uses the case to show that there is a "presumption against shifting harms." See also Malm (1989), Thomson (1991), and Sartorio (2008). In fact, many philosophers claim that it is not even permissible for you to redirect the trolley if the person on the track ahead is *you*. In the debate concerning killing in self-defense, this variation is often used to argue that certain types of non-threatening individuals may not be killed even to save one's own life. In addition to Thomson (1991), see Otsuka (1994), McMahan (2002), and Frowe (2014, 37f and 47).

[4] Consider, for example, Kamm's (1996, 2007) Principle of Permissible Harm. Roughly, it says that a killing can be permissible if it is caused by an act that is (or is "the non-causal flip-side" of) an act of securing a greater good. This principle is silent on cases in which the good that leads to the harm in this way is *not* greater.

case, Trolley K/LD; and generally, many of them are explicit that you may not kill to prevent killing.

This second case contrasts, not killing and letting die but killing performed by you and killing performed by another party. Here is a third example, illustrating a contrast that is still more subtle:

> *Trolley YK/YK* ("you kill vs. you kill")
>
> An out-of-control trolley is running along a track, where it will eventually kill one person. The trolley has been set in motion by a villain in order to kill this person. This villain is you. But you still have time to reconsider. You could use a switch to redirect it onto a sidetrack where it would kill one other person.

Some deontologists find it impermissible to redirect the trolley even in this case, and I am once again inclined to agree with their judgment.[5] In the previous cases, we thought it wrong to sacrifice the person on the sidetrack for the sake of someone else. Surely the person on the side-track does not become any more liable to be killed simply because of your immoral attack on another person.

All three cases illustrate deontic restrictions. They all indicate that, as far as certain intuitive judgments go, we are prohibited from performing an act of killing a person although performing it would prevent the death, or even the killing, or even *our* killing of another. These cases raise questions. Obviously, we should ask for a vindicating explanation of these restrictions. Further, we should prefer a unifying explanation, all else equal. Of course, it *could* turn out that the reasons why you may not redirect in Trolley K/LD and in Trolley K/K are independent. But it should not be our first hypothesis. Also, it is reasonable to hope that the explanation covers Trolley YK/YK.

No simple appeal to the distinction of doing and allowing will cover all of these cases. After all, the latter two cases offer choices between killings and killings. In order for the doctrine of doing and allowing to provide a basis for a unifying account, we would have to make some further claims. First, to cover Trolley K/K it is sometimes claimed that the doctrine of doing and allowing applies in an *agent-relative* way, so that it is your own doing in particular that should be of special concern to you. This would not cover the third case, which is a choice between killings performed by you. But as many authors (Hanna, 2015a; Kamm, 2007; Persson, 2013; Thomson, 1985) note, your *present* choice in Trolley YK/YK does involve a kind of doing/allowing asymmetry. Strictly speaking, this present choice is between killing and allowing a person to be killed as a result of an earlier act of yours. Accordingly, we could propose to apply the doctrine in an *agent-and-present-relative* way. We could suggest that it is your own *present* doing in particular that must be of special concern to you.

But adding these claims to the doctrine of doing and allowing is problematic. The main issue is finding a rationale (cf. Hanna, 2015a; Persson, 2013). Agent-relative restrictions have their critics (Kagan, 1989; Scheffler, 1982), but compared to *agent-and-present-relative* restrictions they seem easy to digest. It is not such a strange idea that you have a special reason to concern yourself with future outcomes if they result from your activities. But why do you have special reason to

[5] For example, Thomson (1985, p. 1414) argues that if we may not kill one person in order to save several others, or to keep a villain from killing these others, then we may not do so in order to prevent killings that we ourselves have initiated earlier. It seems safe to say that Thomson would *a fortiori* endorse this claim when we would prevent only one killing we have initiated, not several. See also Hanna (2015b, p. 87, note 54), who says: "Few people believe that it would be permissible for an agent to kill one person in order to prevent himself from having killed two." Again, the same would seem to go, *a fortiori*, for killing one to prevent oneself from having killed only one.

concern yourself with future outcomes only if they are due to your *present* activity? This is a serious, unresolved issue.

Here is a final desideratum on an adequate explanation. Importantly, whatever the distinction between killing and letting die (and the related contrasts) amounts to, its moral relevance seems to be restricted. A number of authors (e.g., Malm, 1989) observe that it matters whether killing and letting die are options that conflict within a single decision, or whether they are contrasted as options in two separate decision situations. To see this, consider this pair of contrasting examples:

> *Trolley LD/LL* ("letting die vs. letting live")
>
> An out-of-control trolley is running along a track, where it will eventually kill one person. You can use a switch to redirect it onto an empty sidetrack.
>
> *Trolley K/LL*
>
> An out-of-control trolley is running along an empty track, not threatening anybody. You can use a switch to redirect it onto a sidetrack where it would eventually kill one person.

Quite clearly, the choice that leads to the death of the one person is morally impermissible in both cases. Famously, authors like Rachels (1975) and Tooley (1980) conclude from such comparisons that, in general, there is no moral difference between killing and letting die after all.

This complication extends even further. There are single decision situations in which there seems to be no clear intuitive difference between killing and letting die. Here is one such example:

> *Trolley K/LD/LL*
>
> An out-of-control trolley is running along a track, where it will eventually kill one person. You can use a switch to redirect it onto either one of two side-tracks. On side-track one, it will kill one other person. On side-track two, it will not kill anybody.

Here, I would think that there is not much of a difference, morally speaking, between a choice of letting die (by not using the switch) and killing (by directing the trolley onto side-track one). The general lesson seems to be that whenever there is an option in which all affected parties survive, it does not matter much whether you unnecessarily choose to kill a person or to let a person die.

Now, we should not follow Rachels and Tooley and conclude that the difference between killing and letting die *never* matters. As has often been observed, this is unwarranted. Furthermore, let me note that I am not sure that there is no moral difference *at all* between killing and letting die even in these last cases. But I do accept that in these cases, killing and letting die are equally morally *impermissible*. This is different from cases like Trolley K/LD, where letting die, unlike killing, seemed permissible. A correct account of deontic restrictions will have to get these complications right.

Before I present my proposal, let me be clear about what I think is the most serious challenge for a vindicatory account of restrictions on killing. This *Symmetry Challenge*, as we may call it, has been forcefully articulated by, *inter alia*, Shelly Kagan. He summarizes it as follows: "Obviously, then, if there is to be a constraint against doing harm, the moderate must point to something special about doing harm, as opposed to merely allowing harm. [...] [She] must demonstrate

that there is a morally important difference between doing harm and allowing it" (Kagan, 1989, p. 111). This is, in fact, the most natural question to ask about our cases: What is the *difference* between redirecting the trolley and not redirecting it, that makes the former more objectionable than the latter? This question is most naturally addressed by looking for some special element of harm, or some special attitude or relation to the victim etc., that makes killing worse. As far as I am concerned, this search has not produced a convincing answer (although I cannot give a discussion of the literature here).

With this Symmetry Challenge before us, I can highlight what is distinctive of the following proposal. I suggest that we do not need to find a difference that makes redirecting the trolley worse than the alternative. Redirecting the trolley is prohibited, *not* because something makes it worse, but rather because there is *no* morally relevant difference that makes it *better* than the alternative. To be morally permissible, killing would have to be an improvement. So, if we look at the options of redirecting and not redirecting and see only symmetry, *this explains* why redirecting is not allowed.

I claim that a plausible principle of *normative inertia* can motivate this change of perspective.

2 | INTRODUCING NORMATIVE INERTIA

Roughly speaking, principles of normative inertia privilege the path that one is already on. The idea that there are such principles is still underexplored. But there has been some discussion, especially in debates about norms of rational belief. Here, the idea is that once you have adopted a certain belief, there is some normative pressure not to change your mind too readily. In particular, to justify a change of belief, it is not enough to say that the evidence in favor of the revised belief is *no worse*.

Podgorski (2016) offers a discussion of inertia principles for doxastic rationality, which can serve as an entry for us. To begin, he notes that there is broad agreement on two main arguments for such principles. First, they help us to *reduce cognitive costs*, that is, the costs that arise if we change our minds too readily. If a principle raises the normative bar for changes of mind, it directs us to keep these costs low. Second, they help us to *reduce the mental clutter* that would result if we had to keep the original justifying grounds for our beliefs readily in mind to fend off challenges. If a principle makes the beliefs we hold harder to challenge, we have less need to recall those grounds.

Further, Podgorski helpfully distinguishes two different ways to construe normative inertia. All popular proposals (in the literature he is concerned with) adopt a first construal. They suggest that once an option is realized, or in the process of being realized, this generates additional positive support for this option. In the case of doxastic rationality, this means that once we settle on a belief, this adds to the belief's "beliefworthiness." But as Podgorski notes, such views are dubious. For one thing, it is hard to see how the cognitive advantages that they appeal to can be reasons "of the right kind," as they are often called. In the case of doxastic rationality, for example, these views are in tension with the idea that only evidence positively contributes to beliefworthiness. For another, these views must allow that one can "bootstrap" into existence reasons for a belief by adopting it.

Podgorski suggests that a better construal of normative inertia is *dynamic*. This second construal distinguishes between "external" reasons to start, continue, or re-open processes of deliberation, and "internal" reasons that matter within that deliberation, by supporting the

options one deliberates about.[6] Podgorski's core idea is that inertia affects the external or process-related reasons. So once one has a certain belief, one is not always permitted to deliberate on it anew, even if one's current evidence would not suffice to justify it if one *were* to deliberate about the matter.

This second construal avoids the problems of the first one. However, as Podgorski admits, it is also weaker. It entails that once we *do* start to reconsider (even though we may lack sufficient reason to do so), any normative privilege for our given belief vanishes. We must deliberate as if it was the first time. This, as Podgorski admits, does not capture the full strength of our intuitions about inertia. In addition, I have a further concern about this view, not mentioned by Podgorski. As I will argue in more detail below, the requirement to keep our eyes on a separate class of external, process-related reasons would precisely conflict with the underlying aim to reduce cognitive costs.

Anyway, I think that these two broad ways to understand inertia do not exhaust our options. There is a third construal, not discussed by Podgorski. Like the first approach, it is "static" and eschews the invocation of a separate class of process-related reasons. Furthermore, it avoids the weakness of Podgorski's dynamic view. At the same time, unlike that first approach, it avoids the suspicious claim that the fact that a certain option has been taken could somehow add to its merits.

This third construal distinguishes between the *reasons* that justify an option, and the *standard* that determines when those reasons are *of sufficient strength to provide justification*. This distinction is familiar in the domain of epistemic and doxastic rationality. As many authors have stressed (e.g., Harman, 2002; Schroeder, 2021), if the evidence in favor of p is only slightly stronger than the evidence against p (i.e., in favor of $\sim p$), this does not mean that one has sufficient justification for believing that p. Thus, the *relative* strength of the evidence does not by itself determine when a belief enjoys sufficient evidential support. As Schroeder describes it: "[T]he problem of near ties shows that we need some independent conception of what makes evidence sufficient—we cannot simply read it off as the product of its weighing against competing evidence" (Schroeder, 2021, p. 141).

If we accept this distinction between the reasons that carry the justifying weight, and a standard of sufficiency that determines how weighty those reasons have to be, we may suggest that normative inertia affects this standard. For doxastic rationality, the idea would be that once you have adopted a belief, the standard that determines sufficiency of evidential support for this belief is less demanding than the standard of sufficiency of evidential support for its doxastic alternatives.

Interestingly enough, this third construal has been proposed in the area that is our main concern here—that is, moral philosophy. Namely, Malm (1989) defends the following principle:

> **Malm's Principle**
>
> If you change who lives and dies (but not how many) and all else is equal, you need a good reason for doing so; by contrast, *not* changing who lives and dies is justified as long as there is no good reason for changing who lives and dies.[7]

[6] This distinction has been described earlier, for example, by Owens (2000) and Hieronymi (2013). The terminology of "internal" and "external" is from Schroeder (2021, 134 ff).

[7] What I call Malm's Principle is a condensed version of the theses in Malm (1989, 248 f).

As Malm's discussion makes clear, by a "good" reason for some option she means a reason that is not matched by an equivalent reason supporting the alternative option. So on her view, changing who lives or dies is not sufficiently justified if the reasons to do so are merely equivalent to the reasons not to do so. By contrast, not changing who lives and dies *is* justified if the reasons are equivalent. Thus, even if the support by reasons is symmetric, there is *an asymmetric standard for the strength of reasons required for moral justification*. For changes, being equally good is not good enough.

So Malm's Principle is an instance of the third construal of inertia. This is interesting, because in the domain of practical rationality, the need for a separate standard of justification is not often recognized. Standard views simply say that the practical reasons in favor of some option are sufficient iff they are no less weighty than those in favor of the alternatives. This leaves no room for a standard beyond relative weight. The present proposal differs. Roughly, it says: It matters whether the option you choose is to let things take their course, or whether you choose to change the course of events. If you do the latter, the reasons in favor of doing so are sufficient only if they are stronger than those in favor of the *status quo*. So, briefly, if the option you consider is not already in the making (so to speak), the standard of justification is asymmetric. You need *stronger* reasons.

In this respect, Malm's principle diverges from the orthodoxy. However, this by itself seems to me to be not a very major problem. It is not as though the idea of a standard of sufficiency for practical reasons (a standard which is not automatically determined by the relative preponderance of reasons) has received much discussion and has been universally judged to be problematic. Rather, it seems that few authors have seen a need to introduce any such standard. But if it turns out that the idea is serviceable in formulating a plausible principle of normative inertia, why not? To put it more carefully, if the argument that will be presented in favor of normative inertia below are convincing, and if the introduction of a separate standard is the best account of such inertia, that by itself should make us more willing to countenance the idea that there is such a standard.

It should be added that Malm's idea of an asymmetrical standard is not in any serious kind of conflict with the orthodox view. This can be put in precise terms: The idea that normative inertia works by way of an asymmetric standard of sufficiency is *conservative*, in the sense that any choice that is rationally permissible according to it is also permissible according to the orthodox view. If an option has *better* reasons in its favor of than some alternative, Malm's view agrees with the orthodoxy that, out of the two, the former option must be chosen. The only difference between the views arises in cases of symmetrical support. Where one option is already in the making but an alternative enjoys equally strong support by reasons, the orthodox view permits both switching and sticking to the present path, whereas Malm's normative inertia view permits only the latter. Thus, the set of choices that are permissible according to the normative inertia view is always a (proper or improper) subset of the set of choices that are permissible on the orthodox view. We can put this by saying that Malm's normative inertia view is a *conservative strengthening* of the orthodox view.

Alas, there is an important problem for Malm's proposal: She does not succeed in providing a convincing rationale or justification for her principle. Basically, she makes two points. First, she claims that if morality gave us permission to change who lives and dies as long as doing so is no worse, it would give us license to exercise arbitrary control over others. This is not convincing. To say that we should not exercise arbitrary control over who lives and dies pretty much comes to the same thing as saying that we should not change who lives or dies without a good reason. And that is the very claim we want an argument *for*, so this justification begs the question. Her second point is that, if morality gave us the relevant kind of control, we would be permitted to change who lives and dies for all sorts of reasons, some of which may seem objectionable. (She

mentions the example of aesthetic preferences.) But this is a *non sequitur*. After all, the claim that Malm opposes merely says that you need no good reason to change who lives or dies. But that does not entail that you may do it *for* any reason whatsoever. Her opponent can maintain that certain reasons are ruled out.

In the next section, I will suggest that there is a better way to defend Malm's principle.[8] I propose an inertia principle that is not specifically about morality, but about practical rationality. In Section 5, I argue that this verdict can be coupled with a further premise to yield a moral verdict.

3 | NORMATIVE INERTIA, PRACTICAL RATIONALITY, AND DELIBERATION COSTS

The principle I am about to propose says, very roughly, that if you interfere in a certain process with the aim to promote certain outcomes rather than others, this is rational only if this change goes in the direction of outcomes that are strictly better. To motivate this, take the following case:

> *Safe Lottery*
>
> There are *n* 100$ bills in front of Joel. He will receive one of the bills as a gift, selected by a randomizing mechanism. All bills are genuine and in good shape. Just before the lottery starts, Joel points at one bill and asks: "Not that one, please. Can it be replaced? Or can we have a randomizing mechanism select only among the other *n-1* bills? Thank you."

Clearly, this request is odd. For one thing, of course, it seems objectionably ungrateful. But there is more. Manners aside, we would also wonder about Joel's reasons. (In fact, the impression that Joel is ungrateful partly rests on the judgment that there is *insufficient reason* for being so selective.) We would want to ask him: "What is wrong with that particular bill?" This is further confirmed by the fact that we would be tempted to speculate about hidden reasons: Maybe Joel thinks he has spotted a defect in the one bill? Or maybe he enjoys being picky for its own sake? If there is no such hidden motive and Joel sincerely accepts that keeping the first bill would be no worse, and if he nonetheless keeps requesting that the bill be replaced or taken out, he will clearly seem irrational.

A brief note about my use of the term "rational." In some circles, "rational" is tied to self-interest, and used as a synonym of "prudentially rational." I have in mind a familiar usage that is broader and non-substantive. I suggest that if Joel knows that the bill he points to is just as good as the others, his preference is irrational in the sense in which *akrasia*, or a failure to take necessary means to one's ends, or (on some views) having circular preferences, are irrational.[9] (Also, when I say that an action is "rational," or "justified," I mean that rationality *permits* it, not that it *requires* it.)

So, Joel's request seems irrational. But why, exactly? I will argue that the request is needlessly *costly*, and that it violates a normative inertia principle that serves to rule out such wastefulness. Clearly, this is a version of one standard argument for normative inertia, as described in Section 3.

[8] Sartorio (2008) explicitly refrains from trying to justify moral inertia.

[9] The principles of rationality that are violated by these phenomena are sometimes grouped together as "structural," but I doubt that this is adequate. For discussion, see Kiesewetter (2017), Lord (2018), and Henning (2018), inter alia.

To begin, you might think that it is surprising that Joel's request seems irrational. After all, at first glance it seems that the request does not lower the expected value of the lottery (100$). So what is the problem? But on closer inspection, this judgment can be seen to ignore some small factors. In addition to the monetary payoff, there is what economists and psychologists call *deliberation costs*—that is, the time and effort involved in weighing alternatives and arriving at decisions.

It is in this respect that Joel's behavior is costly. He is bringing in new options and generating a need for further decision-making, without a prospect of gain to offset these costs. I suggest that principles of rationality should guide us to avoid such wastefulness. I claim that the orthodox rule to maximize expected value will not do this by itself, and that an inertia principle must be added.

However, this is far from obvious. After all, taking deliberation costs into account, it seems that following the rule to maximize expected value *would* have prevented Joel from making his request. All he would have to do is to apply this rule in a sophisticated way—that is, to anticipate deliberation costs and fold them into the calculation of expected value. The original lottery will occur even without any further deliberation, so if Joel refrains from bringing in more options and choosing among them, no costs need to be subtracted and the expected value will be the full 100$. However, any modified lottery will be held, if at all, only after Joel goes to the length of thinking about which bill(s) to replace etc., so any such outcome will be associated with an additional cost δ, making the expected value of any such option 100$-$\delta$. The orthodox rule gives the right result.

Alas, the literature on deliberation costs gives us good reasons to say that this sophisticated proposal is, in fact, *too* sophisticated. There are two main reasons (see, e.g., Conlisk, 1996). First, if agents consciously adopt the aim to reduce deliberation costs, this will be *self-defeating*. (Folding in deliberation costs adds to the difficulty of a decision problem; so a practice of doing so increases deliberation costs.) Second, it leads to a *regress*. (Before we begin to deliberate about some issue, we have to calculate the costs of that deliberation; but before we do *that*, we need to ask whether this *prior* bit of deliberation is worth the cost associated with *it*, etc.) The lesson to be drawn is that, precisely *because* rational beings should follow rules that reduce deliberation costs, these rules should not make this aim part of their *content* (see Weatherson, n.d.). The aim of reducing deliberation costs is more effectively pursued from within a naïve perspective—a perspective that judges outcomes in abstraction from the costs of deliberation. Thus, rules that guide us should be naïve as well. They should appeal to the value of outcomes and prospects only, in abstraction from deliberation costs.

The natural proposal at this point is the one I am about to make: Rational agents should take care to deliberate only if the costs of deliberation will be offset by some value that can be gained. At the same time, they should do so from a naïve point of view which focuses only on the value of the outcomes, not on deliberation costs. Accordingly, the value that is required to offset these costs will appear, from the naïve perspective that ignores these costs, as a net *surplus*, that is, an *improvement*. In other words, rational agents should deliberate only if better outcomes are in store.

So suppose Joel adopts a naïve perspective and ignores deliberation costs. He keeps his eyes firmly on the outcomes and their value. Now, if he merely follows the rule to maximize expected value, this will not be sufficient to keep him from spending excessive deliberation costs. This rule leaves him with a great range of options which equally maximize expected value: the original lottery, but also the various alternative lotteries that would result from replacing bills, etc. Since this rule leaves Joel with a choice set that features many alternatives, he will have (much) more deciding to do. By contrast, my proposal will leave him with a singleton choice set and no further work to do.

But which precise construal of inertia is the right one? I think the best articulation is the third construal of inertia, as briefly described above. That is, the most plausible account is that once

a lottery is set up, this creates an asymmetric standard of justification, so that Joel is justified in opting for a change only if such a change would be for the better. By contrast, I am less attracted to the first construal, according to which the fact that one lottery is already in the making somehow makes *it* better. As in the case of doxastic rationality, such a view suggests a dubious kind of bootstrapping.

Likewise, I avoid the second construal, proposed for the doxastic realm by Podgorski (see above). On such a view, the fact that the lottery is already set up and about to be held is merely a reason not to *consider* alternatives. I am dissatisfied with this, and for the reasons I have alluded to earlier. First, it is too weak. It allows that once Joel *has* begun to consider alternatives, it is perfectly okay for him to say: "Alright, I admit that the bill is fine, so I shouldn't have opened the issue of other lotteries. But now that I *have* started thinking about it, can you *please* replace it?" Secondly, this dynamic construal requires us to take reasons for and against engaging in deliberation into account. But if we must keep an eye on these further reasons, this will increase deliberation costs.

This completes my argument. Argument for what? I now give a precise formulation of my principle. To allow for precision, I will adopt a decision-theoretic framework, assuming that we assess an option by the sum of the values of its possible outcomes, weighed by the probabilities of these outcomes conditional on the option being taken. My idea is then to introduce an additional comparison. We should not only compare the expected value of our options. Also, we must ask *what would be likely to happen if we were to refrain from making any (further) choice*. (This element is motivated by the pragmatic aim. Since the aim is to reduce deliberation costs, the benchmark is what happens *without* further deliberation.) So we draw on unconditional *ex ante* probabilities regarding what will happen. These probabilities may reflect our existing knowledge of circumstantial facts, natural laws, considerations of normality, etc. But they do not include the expectation that we make a decision.

My principle says that if an act changes the way things are likely to go, some of the outcomes that are made likelier must be better than some of those that are made less likely. Note that this is a very weak version of an inertia principle. It does not say that the change must be for the better *tout court*, maybe in the sense of being strictly better in expectation. Again, it merely says that it is a necessary condition of rationality that the change makes at least *some* better outcome(s) more likely.

Here is some terminology: Say that act H *promotes* outcome B, given a body of background knowledge K, iff $P(B|H\wedge K) > P(B|K)$.[10] Here, P is a probability function measuring the agent's rational credence *ex ante*. Promotion, in my technical sense, is measured against the "default," that is, against the probability of what is likely to happen, given only the background knowledge and not conditional on any choice. Further, say that for mutually exclusive outcomes A and B, act H promotes B *at the expense of* A iff: $P(B|H\wedge K) > P(B|K)$ and $P(A\vee B|H\wedge K) \leq P(A\vee B|K)$. Here, the *increase* in probability of B due to H comes with a *decrease* in the probability of A. H moves probability mass to B and away from A.—The principle I propose can now be stated as follows:

Rational Inertia

If an option H promotes some outcomes at the expense of some other outcomes, choosing H is rationally justified only if at least one of the former is better than at least one of the latter.

[10] In full strictness, I would have to take precautions to rule out cases where H merely "brings news" about the likelihood of an outcome, without causally contributing to it. This, admittedly, would have been easier if I had chosen causal decision theory as an expository framework (see, e.g., Lewis 1981).

So take a decision situation with mutually exclusive and jointly exhaustive outcomes *A*, *B*, and *C*, each of which can be promoted. Suppose that *H* promotes *A*, and only *A*, at the expense of *C*, and only *C*. (This means that the increase in likelihood of *A* comes with a decrease in the likelihood of *C*. The likelihood of *B*, however, is not decreased by *H*. That is to say, by promoting *A*, *H* also promotes $A \vee B$; but *H* does not promote $A \vee C$.) In such a case, the intuitive idea is that the action *H* discriminates between the outcomes *A* and *C* in a way that is justified only if *A* is better than *C*.

To see the principle at work, suppose Joel is first offered a choice between two lotteries, one with and one without the "problematic" bill. Joel must choose; unless he does so, no lottery is held. In this example, my principle allows that choosing either lottery is rationally justified, given that neither is worse than the other one. Choosing one lottery does raise the probability of certain outcomes (i.e., the outcomes of that lottery), but it does not promote them *at the expense* of the outcomes of the other lottery (which are not made less likely compared to the default expectation, so that the probability of *one or the other* set of outcomes is also raised). So there is no asymmetric requirement of justification, and the principle does not leave Joel in a Buridan's-ass predicament.

By contrast, once a lottery *has been* selected (by Joel, or in some other way) and will be held unless Joel interferes, things are different. Now a request for change will raise the probability of certain outcomes (some of those of the new lottery) at the expense of certain other outcomes (some of those of the initial one). My rule says this is irrational, because none of the outcomes that are made more likely are better than the outcomes that were more likely *ex ante*, without a change.

The same underlying thought can also be phrased by adopting terms that were made popular by Ullmann-Margalit and Morgenbesser (1977): If you decide against something that was going to happen by default, then this selection cannot count as mere *picking*. By steering against the anticipated course of events, you count as *choosing* the alternative over this course of events. And choosing, as opposed to mere picking, does require that the selected alternative is better (see Ullmann-Margalit & Morgenbesser, 1977).

Here is a final thing to note. Admittedly, my principle is unorthodox in that it suggests an asymmetric requirement for justification. As I said, this is uncommon in discussions of practical reason. Typically, it is assumed that if options are supported by equally weighty reasons, both (or neither) are justified. In terms of expected value, the assumption is that if two options are equal (or on a par etc.) with regard to their expected value, both (or neither) are justified. Still, as in the case of Malm (see above) my principle is not as revisionary as it may seem. This is because it is *compatible* with the orthodox rule that rational agents maximize expected value (or, alternatively, with the view that they maximize support by reasons). It never leads us to choose options with non-maximal expected value (or reasons-support). So we can keep the orthodox rule and add my view as a further condition. As in Malm's case, this is a *conservative strengthening* of the orthodox rule.

This concludes my discussion of the above rule of Rational Inertia. Of course, much more would have to be said by way of further defense. But hopefully I have convinced you that it merits serious consideration. Next, I will try to explain how this principle can have moral ramifications.

4 | A MORAL PREMISE

We can apply Rational Inertia to our first three trolley cases. In these cases, diverting the trolley is changing the course of events that is expected if no further decision is made. It *promotes* a certain outcome (the survival of the person on the main track). And it does so *at the expense* of a different outcome (the survival of the person on the sidetrack). (Technically, it raises the probability of the former without raising the probability of *one or the other* outcome occurring.) According to the

principle, then, redirecting the trolley is a rational thing to do only if the outcome in which the one on the sidetrack dies is strictly better than the outcome in which the one on the main track dies.

Again, so far this is merely a verdict concerning the rationality of redirecting the trolley, not about the morality of doing so. So, for example, an act utilitarian might well agree with us. She might insist that it is morally permissible to redirect, but grant that since doing so lacks a point, it is irrational.

So how can we get moral mileage out of my principle? We can do so by proposing a further premise which makes the link. I begin by introducing the thought behind this premise in an intuitive way. Here is what I take to be a plausible idea: We should treat persons as equals, in the sense that our actions must be justifiable in a way that is compatible with the equal moral importance of all persons. In other words: If an act is such that a rational person would do it only if she took an unequal view of some persons and the value of their lives, then this act is, *pro tanto*, impermissible.

If something like this idea can be defended, it supplies the needed link between rationality and morality. Here, in a nutshell, is how: Again, according to Rational Inertia, diverting the trolley is a rational thing to do only on the assumption that the life of the person on the sidetrack is less valuable, so sending the trolley toward her strictly improves the prospects. Thus, any candidate for a sufficient justification of the act involves the claim that this person's survival is of lesser value. So by redirecting, you would treat her as inferior. This, in fact, strikes me as a natural diagnosis of the wrong involved in redirecting. Something like this diagnosis is given by Thomson (1991), who remarks that by redirecting in Trolley K/LD, you would "make [the person on the sidetrack] be a substitute victim" (p. 290). You would treat her as if she were more apt for the victim's role.

But how can the idea just floated be defended? I will now explain that it can be given a natural defense in a contractualist framework.[11] After all, the idea expresses a moral constraint on the reasons that justify an action. And the contractualist criterion of moral permissibility is precisely a general constraint on the available justifications of an action. This makes it a natural starting place.

Contractualism (as formulated by Scanlon (1982, 1998)) says that an action is impermissible just in case any principle permitting this action could be reasonably rejected by some affected person (who shares the aim of finding mutually acceptable principles). As Scanlon (1998) explains, principles are derived from justifications of actions in terms of reasons; they state "general conclusions about the status of various kinds of reasons for action" (p. 199). So the demand is "justifiability to each person (on grounds that he or she cannot reasonably reject)" (1998, 390 n.).

When can people reasonably reject our justifications? There are different possibilities. Reasonable rejection can be based on how principles would affect our well-being, or our ability to form reliable expectations about one another's conduct. It can also be based on considerations of fairness. But there may be more. Importantly for our context, Niko Kolodny argues that Scanlonian contractualism should allow a further ground for reasonable rejection. Namely, he proposes that contractualism should embrace the view that "being insulted, being judged to be inferior" will support a "recognizable complaint or objection on the part of the victim" (Kolodny, 2011, p. 120).[12]

This recommendation seems intuitively plausible to me. It is also well motivated, since the contractualist idea of justifiability is supposed to capture an important form of equal recognition.

[11] As I said in the introduction, this is a significant point in its own right. Contractualism is often criticized for its alleged inability to account for deontic restrictions; see Brand-Ballard (2004), Kamm (2007), Shaver (2007), and Suikkanen (2020). Two recent attempts at a contractualist defense of restrictions are Woollard (2015) and Mardellat (2020).

[12] In a note, Kolodny (2011, p. 120, n19) reports that Scanlon has endorsed this view in conversation. See also Kolodny (2023) for a comprehensive account of claims against inferiority.

Now, if we recognize this type of complaint, we can give an argument for my above idea, as follows: Suppose some action is such that it is rationally justified only if some person's life is of lesser importance, compared to the life of another. So any set of reasons that can be given for this action will *either* fail to justify the action *at all*, *or* it will entail that the survival of the person in question is of lesser value. Now I suggest that we follow Kolodny and hold that justifications of this latter kind can be rejected, on account of the judgment of inferiority they imply. As a result, there is *no* principle permitting the act satisfies the standard of justifiability to each person (on grounds she cannot reasonably reject). Any such principle will either not show the action to be justified, or show it to be justified only by reasoning that a relevant person can reasonably reject.

Consequently, I think contractualists have the resources to defend the following premise:

> **Inferiority Implication**
>
> If an action would be rationally justified only if saving some person's life had inferior value compared to saving some other person's life, then (*pro tanto*, and barring special reasons that do support a verdict of inferior value) the act is morally impermissible.

The two principles, Rational Inertia and Inferiority Implication, can now be applied to our cases.

5 | APPLYING THE PROPOSAL TO THE TROLLEY CASES

With the aid of our principles, we can argue thus: By Rational Inertia, redirecting in Trolley K/LD is rationally justified only if the death of the person on the side-track is better than the death of the one ahead. So any principle stating a sufficient rational justification for redirecting will imply that the life of the former person has inferior value. But this satisfies the antecedent of the second principle, Inferiority Implication, which therefore entails that redirecting is morally impermissible.

To clarify this argument, I will address two issues, which will prompt me to explain in more detail how the premises are to be understood. The two issues I have in mind are, first, *personal ties*, and second, *small additional goods*. Briefly, the former seem too partial to make a difference between permissible and impermissible killing, and the latter too insignificant. Can my account capture this?

So first, consider variations of our cases which feature differences in personal ties. Suppose that in Trolley K/LD, the person ahead is your friend.[13] So far, I have stipulated that there are no such differences—but, of course, my account should be able to deliver the correct verdicts either way. Now, I suppose most deontologists would say that such differences do not make killing permissible in our cases. And I agree: In general, it is not permissible to kill an innocent stranger in order to save one's friend (or even, say, one's child). But can my account vindicate this verdict?

At first glance, this is not clear. In a sense, you would not count as *irrational* if you were to prioritize your friend. Your choice would not be erratic, like Joel's request for a substitution lottery among bills. So it could be argued that *Rational Inertia* is satisfied even when one of the outcomes is better only *from a friend's perspective*. But if this is true, then the action of redirecting will merely carry the implication that it is better *from a friend's perspective* if the person on the side-track, a

[13] I thank an anonymous reviewer for urging me to say more about these cases.

stranger, dies. And it is certainly not clear that this latter implication is an insulting judgment of inferiority.

In response, I suggest that *Rational Inertia* must be understood in a specific sense. There is, of course, a sense in which it is rational to act on whatever is our strongest preference. In this sense, saving your friend is undoubtedly rational. But there is a more restrictive sense of rationality in which it involves *responsiveness to reasons*. In this latter sense, partiality *can* be rational, of course, but only if the personal tie really counts as providing a reason. In some cases, it can be a mark of rationality to be able to *abstract* from personal involvements and decide on less subjective grounds.

Now, *is* the fact that the person ahead is your friend a reason to redirect in the present case? I suggest that the criterion is to consider others' perspectives. If the personal tie is any reason at all, then others should reasonably admit that it is. But I do not think the personal tie here meets this (modest) standard. The person on the side-track need not accept that there is a reason why she should have to face death, namely that the two other persons happen to be friends. After all, she is friends with many people, too. And just because they do not have their hands on the switch, that does not seem to be any reason at all why the trolley should be heading anywhere in particular. (Some readers may protest that it is a reason, only not a very significant one. If so, then the case will be taken care of when I address the second "issue" I have announced earlier. Bear with me.)

Thus, I think both *Rational Inertia* and *Inferiority Implication* should be understood as invoking betterness of a kind which, at least, does not *only* obtain from the perspective of a significant other. Hence, redirecting would carry the implication that it is actually better, in a less subjective sense, if the person on the side-track dies, just because the two others are friends—an inferiority implication.

It may be instructive at this point to contrast the following sort of example:

> *Lifeboat*
>
> You are the captain of a lifeboat. You can save either of two castaways on the open sea. But you cannot save both, since your boat can carry only one further person.

In this example, the default (i.e., what happens if you make no further decisions) is that both parties die. A choice to save either party will promote the survival of that party only at the expense of this default outcome, not at the expense of the other person's survival. In consequence, there is no asymmetric requirement of justification for rescuing either party. Rescuing either is permissible.[14]

Now consider a variation of *Lifeboat* involving personal ties. Suppose that, again, one of the castaways is your friend. My account allows that in this case, it *is* permissible to save your friend. But that is not because it is better in any relevant sense if your friend survives, but because the situation is *symmetric*, so there is *no requirement of betterness* in the first place. Therefore, one important lesson is that my view does permit partiality in rescue cases, but not in killing versus letting die cases.

Here, then, is the second issue I have announced: What about cases in which there is a small additional good to be realized by redirecting the trolley, a good of an impartial or objective kind?[15] To take an example from Kamm (1993), suppose that in addition to saving the person ahead, redirecting the trolley would also cure somebody's sore throat. Surely, that is an extra good.

[14] Is there a requirement to give an equal chance to each person, by using a lottery? In Henning (2015), I argue that there is no such requirement.

[15] Thanks to Jakob Lohmar for pressing this worry.

And if so, does not my account allow that killing the person on the side-track becomes morally permissible?

Readers may wonder why I single out improvements that are small, or maybe better: not too great. After all, some deontologists would claim that redirecting is impermissible even if it would secure some very weighty additional good (such as, say, saving further lives). But readers who accept this will typically hold some special and controversial views. For example, they may agree with Taurek (1977) that there is no intelligible sense in which the value of the relevant harms and benefits could be "summed up" across persons. Or they may subscribe to Scanlon's version of contractualism, which considers only complaints by individuals taken separately (this is called the *Individualist Restriction* in Parfit (2011), and only complaints that concern how they are affected personally (this is the so-called *Impersonalist Restriction*, see Parfit (2011)). On such views, even additional goods that are comparatively large may fail to make redirecting overall better in the relevant sense.

As I point out in the final section, I myself am inclined towards a view of this controversial kind. But I assume that many readers will think there is a difference to be made. If, by redirecting, you could achieve some very weighty additional good (e.g., save extra lives), they would agree that redirecting is permissible, or even required. So they do not rule out aggregate betterness in general. Yet even these readers may agree with Kamm that smaller additional goods (of the sore-throat variety) are "irrelevant utilities." They should not make the difference between permissible and impermissible killing. How, apart from full-blown anti-aggregationism, can this be accommodated?

There are at least two ways in which my view can do so. First, I could give up my neutrality on the question of value comparability (voiced at the beginning of Section 1). Since we are dealing with small additional values, I could suggest that the lives of different persons are never determinately equally valuable, but rather incommensurable or "on a par." I could then argue that the relevant evaluative comparisons between outcomes in our cases are "sweetening-insensitive."[16]

A second possibility derives from Kamm's (1993) own solution to the problem of "irrelevant utilities." Her idea is that in interpersonal conflicts, objective and subjective evaluations must be reconciled in our overall value judgment.[17] Within limits, these evaluations can diverge: Each party in our cases may legitimately prefer her own survival, while admitting that *objectively* speaking, the survival of the other is no less valuable—or even that one has a small surplus in value, like a cured sore throat. Consequently, even if the result of redirecting can be said to be *objectively* more valuable, it is not *univocally* better in the sense that accommodates all legitimate perspectives that are involved.

Our two premises can then be understood as requiring betterness in a univocal sense. (Recall that on my account, only the option of redirecting the trolley needs to be based on a univocal judgment of betterness. Not redirecting is permissible if it is not the case that redirecting is better.)

This idea—that is, that within limits, others' points of view must be accommodated—is well motivated in the context of a contractualist theory. (It also fits the criterion of acceptability from others' points of view, which I invoked in the case of personal ties.) After all, our aim is justifiability to all. And we should not hope to convince the one on the sidetrack that her survival is a worse outcome, on the grounds that you will fail to save a person *and* cure a sore throat unless she dies.

According to these two responses, redirecting the trolley does not count as relevantly better, if a small additional good is associated with it. I will not discuss which of them is ultimately right.

[16] For a critical discussion of incomparability and sweetening-insensitivity, see Hare (2010).

[17] Kamm (1993) calls this "sobjectivity."

With these questions addressed, I get back to the main thread. Clearly, my argument against redirecting the trolley will also cover the example Trolley K/K. Here, too, redirecting will be a decision against the default and hence carry an inferiority implication. Clearly, this is independent of whether the default will amount to an instance of killing. All that matters is that it *is* the default, so that any decision against it will be committed to an implication of asymmetric preferability.

Next, my account will also account for the deontic restriction that is in play in the third trolley case, Trolley YK/YK. This may not be obvious. After all, it contrasts a killing which is chosen by you with another killing that is *also* chosen by you. How can we explain the difference?

Start with your first decision, that is, your choice to set the trolley in motion and direct it towards the person ahead. That first choice was morally impermissible, because (we assume) you had the alternative to not set it in motion at all. Now, once the trolley is on its way due to this decision of yours, the default expectation has changed. Henceforth, the way things are expected to go if no further deliberation occurs is that the person ahead dies. In consequence, *redirecting* the trolley is an option that promotes one outcome at the expense of the other; so it is subject to Rational Inertia. That is to say, redirecting is subject to the rational requirement that the outcome of doing so must be strictly preferable to the outcome of not doing so. And as much as the person ahead has reason to object to your previous choice, she is hardly in a position to claim that it is *better* if the person on the sidetrack is killed instead of her. Nor does the person on the sidetrack have to accept this.

Finally, my view vindicates my concession to authors like Rachels and Tooley on their cases (i.e., on the analogous trolley cases I have offered). In both Trolley LD/LL and K/LL, the choice against letting live is wrong for the obvious reason—it leads to a death that could have been avoided at no cost. Of course, in Trolley K/LL, this would also be a choice against the default, and thus carry the objectionable meaning that killing the person on the sidetrack is preferable to not killing anyone. Notably, this is an additional reason to complain, which is absent in Trolley LD/LL. This means that the impermissibility of killing in Trolley K/LL is overdetermined, as it were. So it is not as if there is *no* difference at all between the cases. But it does turn out that both letting die and killing are, when contrasted across cases in this way, both impermissible, and that in both cases there is one sufficient reason for the impermissibility, which is the same. The fact that in Trolley K/LL, there is a further reason, in addition to reasons that already suffice, makes no real difference.

The same argument also applies to the three-options case I have labeled Trolley K/LD/LL.

Note that on this account (unlike many versions of the doctrine of doing and allowing), it is not generally true that being *inactive* is easier to justify. It is refraining from further *deliberation* that is normatively privileged, not necessarily refraining from *action*. If some activity is expected by default, then *it* is easier to justify. Suppose in a variation of Trolley K/LD, you could redirect the trolley, not by using a switch but by holding your breath for 35 s. Surely, the default expectation is that you continue breathing as you normally would. So a decision to stop breathing for 35 s would count as a non-default option, and thus be impermissible (see Sartorio, 2007).

6 | WITHDRAWING AID

To further highlight the explanatory potentials of my view, in this section I apply it to examples of a kind that have seemed puzzling to recent authors, namely cases of withdrawing aid. Consider:

Two Trolleys I

> Two out-of-control trolleys are running along two separate tracks. Each is running towards one person. On the left track, there is a boulder that will stop the trolley before the person on that track is killed. On the right track, there is no boulder. The person on the right track is your friend. You could stop the right trolley and save your friend, but only by moving the boulder from the left track onto the right track, leaving the person on the left to die.

Two Trolleys II

> To your left, an out-of-control trolley is running along a track where it will eventually kill one person. There is a boulder nearby, which you lift onto the track and which will stop the trolley before it kills the person. Seconds later, you realize that to the right, there is a second track and a second out-of-control trolley running toward your friend. You could remove the boulder to block the right track instead, leaving the person on the left to die.

To be upfront, I think it is permissible to save your friend in the second case but not in the first. At the same time, however, the cases are obviously very similar. In both cases, the cost of saving your friend is exactly the same—the death of the stranger on the other track. And, strikingly, even the mode of your agency is very similar. Your acts exemplify what Foot (1967) calls *enabling* harm—that is, the removal of a barrier that would have brought a sequence leading to the harm to a stop.

My intuitive judgment (i.e., that there is nonetheless a difference in permissibility between the cases) is shared by other authors. A much-discussed analysis is given by McMahan (1993): The cases illustrate a moral difference between *(a)* terminating aid which you have not provided yourself, or which is already operative and self-sustaining in that it requires no further efforts from you, and *(b)* withdrawing aid which you have provided yourself and which is either not self-supporting or not yet operative. The former type, McMahan claims, is morally equivalent to killing. It is illustrated by *Two Trolleys I*, where the aid you would terminate was not provided by you. For McMahan, moving the boulder in this case amounts to killing, and is wrong. The second type of case is considered equivalent to letting die. It is illustrated by *Two Trolleys II*. Here, the aid (i.e., the boulder on the first track) was provided by you and it is not yet operative, that is, it has not yet had any effect on the actual sequence of events. Here, removing the boulder and using it to save the other person amounts to letting one person die rather than the other, which is considered permissible.

McMahan's article has convinced some that this intuitive difference is real. But the question is, of course, *why* the two kinds of enabling should differ morally. As Hanna (2015b, p. 83) notes, the explanation seems comparatively easy in cases where the aid is not self-sustaining. As the aid requires more activities, terminating it is a matter of not being active, that is, of allowing harm. So whatever accounts for the moral difference between doing and allowing can be expected to apply.

However, this strategy does not work in cases where the aid would not require any further efforts (i.e., is self-sustaining) but is simply not yet operative (in the sense that is has not yet had any effect on the actual course of events). As Hanna observes, in such a case the relevant difference seems to boil down to the question of whether the presence of the aid is due to your past activity.

Consequently, Hanna argues, it seems that the intuitive verdict must rest on principles concerning a special moral relevance of one's own past behavior. But these are dubious (Hanna, 2015a, 2015b).

My case *Two Trolleys II* is designed to be one of these critical cases; the aid is self-sustaining but not yet operative. I now argue that my account can support the intuitions about these cases.

To begin, here is why *Two Trolleys II* presents a difficulty for my view. Initially, the case is like *Lifeboat*. Before you put the boulder on the tracks, the default expectation is that neither person is saved. So there is symmetry, and both options of saving someone are permissible. So, as I explained above (in the context of *Lifeboat*), there is initially no objection to saving your friend. However, once you *have* put the boulder on the first track, a choice of saving your friend would go against the default. It would promote the survival of your friend at the expense of one particular alternative, that is, the survival of the stranger. By Rational Inertia and Inferiority Implication, this is impermissible.

Still, I think that my proposal can handle these cases. Note that *Two Trolleys II*, just like the corresponding case in McMahan (1993, p. 262) (called *Burning Building 2*) has a special dramaturgical feature. You, the protagonist, receive new information at a crucial point in time. In contrast to *Two Trolleys I*, you realize that there is a second track and a second trolley only after your first decision.

I find that this dramaturgical feature is quite important to my intuitions about the cases. If you had decided to save the stranger rather than your friend in full knowledge, and if the boulder was now in place, I would no longer see an important difference to *Two Trolleys I*. (Of course, your friend could protest that by failing to exploit a permissible way to be partial, you have revealed yourself to be not a true friend. But once the boulder has been placed on the other track, that damage is done, and changing again would be objectionable in its own way.) So what makes *Two Trolleys II* different is the fact that you find out about the second track and your friend only later.

The crucial advantage of my approach, with its appeal to Rational Inertia and the pragmatic aim of avoiding excessive deliberation costs, is that it can *explain why* this particular feature matters. After all, any principle of rationality that serves the pragmatic aim of reducing deliberation costs in an adequate way must accommodate the fact that *sometimes*, more deliberation is a good thing, and worth the cost. So any such rule must be appropriately sensitive to special cases where unexpected information that looks like it might possibly make a difference to the decision becomes available.

Thus, given that I have defended Rational Inertia in terms of the goal to reduce deliberation costs, I have an independent reason to conceive of it in such a way that it *is* sensitive in this way.

In what way, exactly? One possibility would be to say that a rule that is properly sensitive in this regard should contain an explicit proviso concerning cases where unexpected and potentially relevant information becomes available. We would then reformulate Rational Inertia accordingly. However, I prefer to keep Rational Inertia in its simpler form and suggest that we understand it as a *defeasible* principle, which is defeated when unexpected, potentially relevant information arrives.

Like I said, given the pragmatic rationale behind Rationale Inertia, this is something we should expect. And once we adopt such an understanding of the rule, it is possible to account for *Two Trolleys II*. After all, *Two Trolleys II* is precisely the type of case which an adequate view will treat as exceptional and as a defeater: unexpected information about a further option becomes available.

So upon the arrival of this new information, the defeasible principle that serves to keep you from too much deliberation is overridden. Hence there is no asymmetric standard of justification—you are justified in promoting either outcome as long as it is not worse than the other. In effect, the

situation now presents itself like *Lifeboat* all over again. You can simply ignore that one of the options was already in the making. It is permissible to save either party, and so to save your friend.

7 | CONCLUSION

Since the first trolley scenario was introduced in Section 1, the reader may have wondered what my account will say about the more well-known cases involving unequal numbers. What if, by redirecting the trolley and killing the one person, you would save not *one* but *five* persons ahead?

At this point, it must be stressed that an account that sheds light on the relevance of the difference between doing harm and allowing harm, like the present one, should not be expected to "solve" the notorious trolley problem. This is because that trolley problem *is designed* to be a contrast between cases that *agree* with regard to the doing-allowing distinction. Thomson (1985) explicitly uses the comparison between the "bystander" case and the "footbridge" case with the aim to show that the original suggestion by Foot (1967), which appeals to the difference between doing and allowing (and between positive and negative duties), does not go to the heart of the matter.

Thus, if readers should agree with the majority opinion that it is permissible to kill the one in the bystander case but not in the footbridge case, it is clear that no account of the doing/allowing asymmetry will explain these verdicts by itself. The present view is no exception. In both examples, killing the one is a decision against the default, and thus permissible only if the survival of the five is a better outcome. Now, if it is indeed better, my principles will permit redirecting in the bystander case. By the same token, they will not explain why it is wrong to kill in the footbridge case. (If the readers in question do think it is wrong to kill here, they will need further explanatory resources.) If, by contrast, we agree with Taurek (1977) that it makes no sense to say that the survival of the five is better, period, *or* if we accept Scanlon's *Individualist* and *Impersonalist Restrictions* (see above), my theory will lead readers to Thomson's (2008) view that it is wrong to save the five in both cases.

Full disclosure: I myself tend towards this latter option. Pardon me for indulging in some self-explanation. In recent work (Henning, 2024), I defend the claim that differences in numbers can matter in rescue scenarios, that is, in scenarios of the *Lifeboat* type. My defense eschews the appeal to "aggregative" betterness. Instead, it relies on the idea that the affected persons have claims to decide for themselves. We should respect these claims by being guided by our beliefs about what they would decide, if we could give each a vote. This will often lead us to save the greater number.[18]

However, already in that work I have indicated that the case of *doing harm* is different. I can now give more explanation: As I have argued here, one of the critical questions in the relevant scenarios turns out to be the question of whether any further decision-making is warranted. Now, in *Lifeboat* cases, there is a univocal judgment, acceptable by all, that it is better to make *some* decision against the default, lest *all* the affected persons die. (There is, as I put it earlier, a convergent interest in a decision being made.) So, by this univocal judgment, there is an improvement in store and further decision-making is warranted. And then, again, the power to make that decision should be shared among the affected persons. By contrast, in trolley cases, it is far from clear that there is such a univocal judgment that further decision-making is warranted in the first place. The person on the side-track will see no such warrant, denying that there is an improvement in store. Unless we make

[18] It has since been pointed out that my account will, in some cases, systematically lead to verdicts that do not prioritize the greater number. See, for instance, Werner (2024) and Tank et al. (n.d.).

the separate "aggregationist" claim that the survival of the five is better (in a way that permits us to discard dissenting evaluations), my view will *not* endorse any further decision-making. And if no further decision-making is warranted at all, no majoritarian decision is warranted either. So although numbers do matter in rescue cases, they may not do so in cases of the bystander-type.

Still, I realize that many readers will prefer a different solution. One way to achieve this would be, as indicated, to accept Parfit's proposals regarding contractualism and to drop the Individualist and Impersonalist Restrictions. One can then argue that redirecting the trolley in the bystander case is permitted on my account. Again, like everyone else who leans towards this view, these readers will need additional resources if they wish to explain why the footbridge case is different.

Time to wrap up: I have proposed a principle of rational decision-making and argued that, when coupled with a moral premise that contractualists should accept, it vindicates a range of intuitive and commonsense judgments about restrictions on killing. There are further examples and complications that need to be considered. But I have covered more than enough for one article.

ACKNOWLEDGMENTS

I am grateful to many people for helpful discussions, especially to Hauke Behrendt, Stephanie Elsen, Jakob Lohmar, Samuel Ulbricht, Konstantin Weber, to audiences in Erlangen-Nürnberg and in Münster, and, last but not least, to helpful suggestions from reviewers.

Open access funding enabled and organized by Projekt DEAL.

REFERENCES

Brand-Ballard, J. (2004). Contractualism and deontic restrictions. *Ethics*, *114*, 269–300. https://doi.org/10.1086/379354.

Conlisk, J. (1996). Why bounded rationality? *Journal of Economic Literature*, *34*, 669–700.

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

Frowe, H. (2014). *Defensive killing: An essay on war and self-defense*. Oxford University Press.

Hanna, J. (2015a). Doing, allowing, and the moral relevance of the past. *Journal of Moral Philosophy*, *11*, 677–698. https://doi.org/10.1163/17455243-4681049.

Hanna, J. (2015b). Enabling harm, doing harm, and undoing one's own behavior. *Ethics*, *126*, 68–90. https://doi.org/10.1086/682190.

Hare, C. (2010). Take the sugar. *Analysis*, *70*, 237–247. https://doi.org/10.1093/analys/anp174.

Harman, G. (2002). Practical aspects of theoretical reasoning. In A. Mele & P. Rawling (Eds.), *The Oxford handbook of rationality*. Oxford University Press.

Henning, T. (2015). From choice to chance? Saving people, fairness, and lotteries. *Philosophical Review*, *124*, 169–206. https://doi.org/10.1215/00318108-2842176.

Henning, T. (2018). *From a rational point of view*. Oxford University Press.

Henning, T. (2024). Aggregation without numbers. *Noûs*, *58*, 755–777. https://doi.org/10.1111/nous.12475

Hieronymi, P. (2013). The use of reasons in thought (and the use of earmarks in arguments). *Ethics*, *124*, 114–127. https://doi.org/10.1086/671402

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

Kamm, F. M. (1993). *Morality, mortality Vol. I*. Oxford University Press.

Kamm, F. M. (1996). *Morality, mortality Vol. II*. Oxford University Press.

Kamm, F. M. (2007). *Intricate ethics: Rights, responsibilities, and permissithroatble harm*. Oxford University Press.

Kiesewetter, B. (2017). *The normativity of rationality*. Oxford University Press.

Kolodny, N. (2011). Scanlon's investigation: The relevance of intent to permissibility. *Analytic Philosophy*, *52*, 100–123. https://doi.org/10.1111/j.2153-960X.2011.00522.x.

Kolodny, N. (2023). *The pecking order. Social hierarchy as a philosophical problem*. Harvard University Press.

Lewis, D. (1981). Causal decision theory. *Australasian Journal of Philosophy*, *59*, 5–30. https://doi.org/10.1080/00048408112340011.

Lord, E. (2018). *The importance of being rational*. Oxford University Press.

Malm, H. M. (1989). Killing, letting die, and simple conflicts. *Philosophy and Public Affairs*, *18*, 238–258.
Mardellat, V. (2020). Contractualism and the paradox of deontology. *Philosophical Studies*, *177*, 3749–3774. https://doi.org/10.1007/s11098-019-01406-w
McMahan, J. (1993). Killing, letting die, and withdrawing aid. *Ethics*, *103*, 250–279. https://doi.org/10.1086/293495
McMahan, J. (1994). Self-defense and the problem of the innocent attacker. *Ethics*, *104*, 252–290. https://doi.org/10.1086/293600.
McMahan, J. (2002). *The ethics of killing: Problems at the margins of life*. Oxford University Press.
Otsuka, M. (1994). Killing the innocent in self-defense. *Philosophy and Public Affairs*, *23*, 74–94. https://doi.org/10.1093/0199243956.003.0005.
Owens, D. (2000). *Reason without freedom*. Routledge.
Parfit, D. (2011). *On what matters, Vols. 1 and 2*. Oxford University Press.
Persson, I. (2013). *From morality to the end of reason: An essay on rights, reasons, and responsibility*. Oxford University Press.
Podgorski, A. (2016). Dynamic conservatism. *Ergo*, *3*, 349–376. https://doi.org/10.3998/ergo.12405314.0003.013
Rachels, J. (1975). Active and passive euthanasia. In S. M. Cahn (Ed.), *Exploring philosophy: An introductory anthology*. Oxford University Press.
Sartorio, C. (2008). Moral inertia. *Philosophical Studies*, *140*, 117–133. https://doi.org/10.1007/s11098-008-9229-x.
Scanlon, T. M. (1982). Contractualism and utilitarianism. In A. K. Sen & B. Williams (Eds.), *Utilitarianism and beyond* (pp. 103–128). Cambridge University Press.
Scanlon, T. M. (1998). *What we owe to each other*. Belknap Press of Harvard University Press.
Scheffler, S. (1982). *The rejection of consequentialism: A philosophical investigation of the considerations underlying rival moral conceptions*. Clarendon Press.
Schroeder, M. (2021). *Reasons first*. Oxford University Press.
Shaver, R. (2007). Contractualism and restrictions. *Philosophical Studies*, *132*, 293–299. https://doi.org/10.1007/s11098-005-2219-3.
Suikkanen, J. (2020). *Contractualism*. Cambridge University Press.
Tank, L., Wendler, N., & Carstensen-Mainka, J. P. (n.d.). Unpublished manuscript.
Taurek, J. (1977). Schould the Numbers Count? *Philosophy & Public Affairs*, *6*, 293–316.
Thomson, J. (1985). The trolley problem. *Yale Law Journal*, *94*, 1395–1415. https://doi.org/10.2307/796133.
Thomson, J. (1991). Self-defense. *Philosophy and Public Affairs*, *20*, 283–310.
Thomson, J. (2008). Turning the trolley. *Philosophy and Public Affairs*, *36*, 359–374. https://doi.org/10.1111/j.1088-4963.2008.00144.x.
Tooley, M. (1980). An irrelevant consideration: Killing versus letting die. In B. Steinbock (Ed.), *Killing and letting die*. Prentice-Hall.
Ullmann-Margalit, E., & Morgenbesser, S. (1977). Picking and choosing. *Social Research*, *44*, 757–785.
Weatherson, B. (n.d.). Deliberation costs. [Unpublished manuscript]. Retrieved July 13, 2025, from http://brian.weatherson.org/Deliberation_Costs.pdf.
Werner, J. (2024). Rescue cases, the majority rule, and the greatest number. *Journal of Ethics and Social Philosophy*, *28*. https://doi.org/10.26556/jesp.v28i3.3691
Woollard, F. (2015). *Doing and allowing harm*. Oxford University Press.

How to cite this article: Henning, T. (2025). Killing, letting die, and normative inertia. *Philosophy and Phenomenological Research*, *111*, 728–749. https://doi.org/10.1111/phpr.70053