

THOMSON'S TROLLEY PROBLEM

Peter A. Graham

NO ONE HAS DONE MORE over the past four decades to draw attention to the importance of, and attempt to solve, a particularly vexing problem in ethics—the Trolley Problem—than Judith Jarvis Thomson. Though the problem is originally due to Philippa Foot, Thomson showed how Foot's simple solution would not do and offered some of her own.¹ No solution is uncontroversial and the problem remains a thorn in the side of non-consequentialist moral theory. Recently, however, Thomson has changed her mind about the problem. She no longer thinks she was right to reject Foot's solution to it. I argue that, though illuminating, Thomson's current take on the Trolley Problem is mistaken. I end with a solution to the problem that I find promising.

In sections 1–3, I present Thomson's version of the Trolley Problem (one involving a twist on Foot's original version) and her various responses to it. In sections 4 and 5, I evaluate her various takes on the problem, including her most recent rejection of the problem. In section 6, I offer a diagnosis of the purported data on the basis of which Thomson has mistakenly come to reject the problem. And in section 7, I present and defend my own preferred solution to the Trolley Problem.

1. THE PROBLEM STATED

Foot's version of the Trolley Problem revolves around pairs of cases like these:

Big Man: An out-of-control trolley—the driver is unconscious—is barreling toward five workmen trapped on the track ahead of it. If nothing stops the trolley, the five will be run over and killed. A big man whose weight would stop the trolley before it reaches the five were he to sacrifice his life by jumping in front of it decides not to. However, a thin man can push the big man into the path of the trolley.

Driver: An out-of-control trolley is barreling toward five track workers

¹ Foot, "The Problem of Abortion and the Doctrine of Double Effect."

who are trapped on the track ahead of it. If the driver does nothing, the five will be run over and killed. The driver cannot stop the trolley, but he can turn it onto a spur of track to the right, on which there is another trapped track worker who would be run over and killed were he to do so.

Foot took it to be obvious that, though the thin man may not push the big man, the driver may turn the trolley. And the problem she took herself to be addressing was why that was the case. After all, in both cases the agent faces the choice of whether to kill one to save five. Her answer was that the cases were importantly asymmetrical: though the thin man lets five die if he does not kill, the driver kills either way. As Thomson puts it: "If the driver fails to turn his trolley, he does not merely let the five track workmen die; he drives his trolley into them, and thereby kills them."² And if, as Foot suggested, the following two claims are true,

(I) killing one is worse than letting five die, and

(II) killing five is worse than killing one,

then the solution to her problem was right at hand. As (I) dictates that pushing the big man is worse than not pushing him, pushing the big man is impermissible. But, as (II) dictates that not turning the trolley is worse than turning it, turning it is permissible. Furthermore, Foot argued, this solution comports with our intuitive conviction that negative moral duties—e.g., the duty *not* to kill—are more stringent than positive moral duties—e.g., the duty *to* rescue.

Thomson argued that Foot's solution fails.³ She suggested we consider:

Bystander: An out-of-control trolley—the driver is unconscious—is barreling toward five track workers trapped on the track ahead of it. A bystander can either (i) do nothing, in which case the five will die, or (ii) flip a switch to the right, diverting the trolley onto a right-hand spur of track away from the five, thereby killing another track worker who is trapped there.

It is no less morally permissible, Thomson claimed, for the bystander to flip the switch than it is for the driver to turn his trolley in *Driver*. Foot's (I) and (II) offer no help in explaining this. Why may the bystander turn the trolley when he, like the thin man, chooses between killing one and letting five die? What is needed is a solution that accommodates the data in all the cases.

² Thomson, "The Trolley Problem," 1397.

³ See Thomson, "The Trolley Problem."

2. THOMSON'S SOLUTIONS TO THE PROBLEM

In "The Trolley Problem," Thomson offered a solution—call this her First Solution—according to which the bystander may flip the switch in *Bystander* because were he to do so (1) he makes what was threatening five come to threaten only one and (2) he does so not by any means that constitute an infringement of any right of the one's.⁴ If the thin man pushes the big man in *Big Man*, by contrast, though (1) would be true, (2) would not: he would make the threat to the five threaten the one by means that do infringe the one's rights, in particular, the right the one has against him that he not push him. But does the bystander not infringe the one's rights in turning the trolley onto him? In one sense, yes, but in another, no. True, the one has a right that the bystander not kill him and the bystander does kill him. But the turning of the trolley does not *in and of itself* violate any of the one's rights. So, though the bystander does indeed infringe the one's right not to be killed, his doing that which saves the five—turning the trolley—does not, in and of itself, infringe any right of the one's.

Thomson's First Solution reflects a certain picture that many proponents of rights-based moral theories find attractive. There is a standing moral requirement to minimize harm wherever possible. However, people's rights prevent them from being permissibly sacrificed to do so. To use Ronald Dworkin's evocative metaphor, rights trump utilities, and on this picture they do so by being *means-blocking barriers*.⁵ Rights act to constrain how we may go about minimizing overall harm by blocking certain causal pathways to that result; only if it is caused *in certain ways* is it permissible to cause harm in order to minimize harm overall. Our moral worth and integrity are in large part constituted by the rights we have and so only by avoiding these barriers do we respect the moral worth and integrity of others.

In *The Realm of Rights*, however, Thomson rejects her First Solution because she now thinks that the bystander's turning of the trolley is indeed an infringement of the one's rights.⁶ Having rejected her First Solution, Thomson then offers a Second Solution: the bystander may turn the trolley because all six workmen—the five and the one—belong to a group such that at some point in the past it was in each member's individual interest that the bystander turn the trolley were *Bystander* ever to arise. As track assignments were determined by

4 Thomson also discusses the Trolley Problem in "Killing, Letting Die, and the Trolley Problem"; however, there she gestures at, but does not explicitly defend, a solution to it.

5 For Dworkin's talk of rights trumping utilities, see Dworkin, "Rights as Trumps."

6 She thinks this because she now accepts a principle from which it follows that turning the trolley infringes one of the one's rights. I discuss this in detail in section 5.

lot, each workman's expected utility, prior to that determination, is greater if the bystander flips the switch should *Bystander* arise than if he does not. And it is in virtue of this, she held, that the bystander may flip the switch.

These are Thomson's solutions to the Trolley Problem. I think her First Solution is on the right track and she veered off it with her Second Solution. I say why below. For now, however, we need to look at her third and final take on the problem. This time she does not offer a solution to the problem; instead, she rejects that there is a problem in need of solving at all.

3. THOMSON'S ABOUT-FACE

Thomson has had a change of heart about *Bystander*. She now thinks the bystander's turning the trolley is impermissible. Her argument begins with:

Three Options: Things are as they are in *Bystander* except that the bystander has a third option: (iii) he can flip the switch to the left, diverting the trolley onto a left-hand spur of track on which he himself is trapped, thereby saving the five but killing himself.

Thomson takes it to be obvious that, though it would be permissible for the bystander to choose either (i) or (iii)—forgoing killing someone, even if it means letting five innocent people die, and voluntarily sacrificing oneself in order to save five innocent people from dying are both certainly permissible—it would *not* be permissible for him to choose (ii). What these reflections show, Thomson maintains, is the truth of the following:

(*) : If *A* wants to do a certain good deed, and can pay what doing it would cost, then—other things being equal—*A* may do that good deed only if *A* pays the cost himself.⁷

And, in particular, it follows from (*), she claims, that a certain *ceteris paribus* principle is true:

Third Principle: *A* must not kill *B* to save five if he can instead kill himself to save them.⁸

All of this, she maintains, supports the impermissibility of option (ii) in *Bystander*.

Thomson suggests that if the bystander would not turn the trolley onto himself and pay the cost of his own life to save the five were he instead in *Three Options*, then “there is no way in which he can decently regard himself as entitled

7 Thomson, “Turning the Trolley,” 365.

8 Thomson, “Turning the Trolley,” 365.

to make someone else pay it [in *Bystander*].”⁹ But this could not really be that which underwrites the wrongness of the bystander’s turning the trolley. It cannot really be thought that, though it is impermissible for the bystander to turn the trolley in *Bystander*, it would have been permissible had he been just a little bit more of a noble sort. In fact, Thomson agrees. Of the bystander who *would* turn the trolley on himself were he in *Three Options* instead of *Bystander*—“the altruistic bystander”—she says:

[He] is not entitled to assume that the one workman is equally altruistic, and would therefore consent to the bystander’s choosing option (ii). Altruism is by hypothesis not morally required of us. . . . Suppose, then, that the bystander knows that the one workman would not consent, and indeed is not morally required to consent, to his choosing option (ii). The bystander has a permissible alternative, namely choosing option (i)—that is, letting the five die. I think it very plausible therefore that there is no way he can justify to himself or to anyone else his choosing option (ii), and thus he cannot decently regard himself as entitled to choose it.¹⁰

What matters here, whether the bystander is altruistic or not, then, is that the one does not, and is not required to, consent to being killed and the bystander has a permissible alternative.

If Thomson is correct here, then it is not morally permissible for the bystander to turn the trolley in *Bystander*. As the permissibility of that option was the basis of her rejection of Foot’s solution to her problem—that of explaining the data in *Big Man* and *Driver*—Foot’s solution to that problem stands. And, as there is no moral difference between *Big Man* and *Bystander* to be explained, Thomson concludes, the Trolley Problem is a “nonproblem.”

4. EVALUATING THOMSON’S ABOUT-FACE

For Thomson, what is crucial to the impermissibility of the bystander’s taking option (ii) in *Bystander* is that the one neither consents nor is morally required to consent to his taking it and also that the bystander has a permissible alternative. Thomson (implicitly) offers a principle:

(&): X may make Y suffer a harm in doing a good deed only if either (a) Y consents (or would consent) to X’s doing it, (b) Y is morally required to consent to X’s doing it, or (c) X has no permissible alternative to doing it.

9 Thomson, “Turning the Trolley,” 366.

10 Thomson, “Turning the Trolley,” 367.

Perhaps we should accept it. "Who are you," you might think, "to make someone else suffer a harm she does not and need not acquiesce to in suffering to achieve an end you need not bring about?"

Should we accept (&)? I think not. To see why, consider:

Swedes: Chang finds himself in the central square of an isolated Swedish mountain village. Against the wall are five innocent teenage boys whom the local sheriff is about to have executed as punishment for a revolt by their parents. Because Chang is a visitor, however, the sheriff offers him the "privilege" of executing one of the five, Sven. If Chang accepts, the sheriff will release the four other boys in celebration of the special occasion. If he declines, then the sheriff will execute all five as he had planned. Chang asks Sven if he consents to his shooting him, but, because the sheriff will kill Sven's sister if he does, Sven refuses.¹¹

Swedes is a counterexample to (&). Chang may make Sven suffer the harm of death in order to do the good deed of saving the four other boys even though none of (&)'s clauses is satisfied: Sven does not, nor is he morally required to, consent to Chang's killing him, and Chang does have a permissible alternative to killing Sven, viz., he may refrain from killing him.

But maybe consent is a red herring. Perhaps what is crucial to the impermissibility of the bystander's turning the trolley onto the one in *Bystander* is that the one would not be required to bring the harm in question upon himself in order to bring about the good in question had he the option of doing so. Perhaps what Thomson is (implicitly) appealing to, then, is:

(&') X may make Y suffer a harm, *h*, in doing a good deed, *g*, only if either (a) Y consents (or would consent) to X's doing so, (b) were Y in a position to bring about *g* at the cost of suffering *h* herself, Y would be morally required to bring about *g* and suffer *h*, or (c) X has no permissible alternative to doing so.

Maybe this will do the trick.

It will not. Consider:

Shepherd: An out-of-control trolley is barreling toward five track workers who are trapped on the track ahead of it. Nothing can stop the trolley before it reaches the five and there are no side spurs of track onto which the trolley can be diverted. A shepherd standing by the side of the track

11 This is a variation on Williams's famous Jim and the Indians case in Smart and Williams, *Utilitarianism*.

can use his crook to pull the five off the track before the trolley reaches them (option (ii')). Unfortunately, if he does this, though he will save the five, the trolley will then run over and kill another track worker trapped a hundred yards behind the five. Were he to not pull the five off the track (option (i)), then though the five would die, the one would not be killed, for the weight of the five dead bodies would halt the trolley well in front of him.

Both (i) and (ii') are clearly permissible. *Shepherd* is thus a counterexample to (&'). Since the shepherd may take option (ii') and he has a permissible alternative, viz., (i), and the one would not be required to remove the five from the track if he could, (&') is false.

These cases cast doubt upon (&) and (&'). What is more, Thomson even grants that it is intuitive that the bystander may turn the trolley in *Bystander*; that is why she needs an argument to toss it overboard. But no argument within the vicinity withstands scrutiny. So, the Trolley Problem is not a nonproblem. It needs a solution.

5. RECONSIDERING THOMSON'S REJECTION OF HER FIRST SOLUTION

Though it does need a solution, we need not look far to find the beginnings of one. Thomson's First Solution contains an important moral insight, which in section 7 I will argue is an integral part of the correct solution. But first I will explain why her Second Solution fails and why her reasons for abandoning her First Solution were bad ones.

Thomson's Second Solution is a nonstarter. According to it, the bystander may turn the trolley in *Bystander* because doing so was in the interest of all six track workers prior to when their track positions were determined. But were that the reason, pushing the big man in *Big Man* would be permissible if the big man himself were also a worker on his lunch break. Pushing the big man in that case would not be permissible, however. Thomson's Second Solution thus fails.¹²

What is more, her reasons for abandoning her First Solution were bad ones. Thomson abandoned her First Solution because she came to endorse:

The Means Principle for Rights (MPR): If (i) *X* has a right against *Y* that *Y* not do β , and (ii) if *Y* does α then *Y* will thereby do β , then *X* has a right against *Y* that *Y* not do α , that right being at least as stringent as *X*'s right against *Y* that *Y* not do β .

12 Kamm raises a similar kind of objection (*Morality, Mortality*, 2:167).

I grant that if MPR is true, then First Solution fails—if MPR is true, the one *does* have a right that the bystander not turn the trolley (because the one has a right that the bystander not kill him, and if he turns the trolley the bystander will thereby kill him). But we should not accept MPR.

Why should we think MPR is true? Thomson offers two reasons. Here is the first:

Does anybody have a [right] against me that I not press my doorbell? Nicholas, for example? What on earth could be reason to think he does? Then we learn that if I press the doorbell, I will thereby kill him. (He and a battery are wired to the doorbell.) ... [MPR] tells us that he does in fact have a [right] that I not press the doorbell. Admittedly, I would be doing him no wrong if it were not the case that by pressing the doorbell I would do him a harm. But that *is* the case. And don't we think he therefore has a [right]—a very stringent [right]—that I not press the doorbell?¹³

This is unpersuasive. What is true is that if I press my doorbell I will thereby infringe Nicholas's right against me that I not kill him. We need not add to this that Nicholas has *another* right against me that I not press my doorbell.

Nothing is gained, and something may well be lost, by supposing Nicholas has a right that I not press my doorbell. If MPR is correct, Nicholas has *two* distinct rights against me, each as stringent as a right not to be killed, both of which would be infringed if I press my doorbell. This is double counting. Often, if *X* has two rights of stringency *S* against *Y* that *Y* would infringe were *Y* to ϕ , the weights of those rights combine, increasing the moral pressure against *Y*'s ϕ -ing above what it would be if *X* had only one such right. But it would be quite odd, as MPR seems to entail, if faced with a choice between causing *X* a harm directly, i.e., not by doing something else, and causing *Y* a similar harm by ϕ -ing, I ought to harm *X*, for in doing so I only infringe one right of *X*'s against me—that I not harm *X*—whereas if I cause *Y* the harm, I infringe two rights of *Y*'s against me—that I not harm *Y* and that I not ϕ —each equal in stringency to that of *X*'s right against me. Now, a proponent of MPR might say that rights connected via MPR do not combine in the way distinct rights do. However, if they do not, then MPR-generated rights, if such exist, are a very special class of rights, different in kind from all others. But, given that, why think there are any such rights? Surely we need not lean on Nicholas's having a right against me that I not press my doorbell to explain why I ought not to press it. We can explain that by appeal to his having a right that I not kill him and that if I press my doorbell I will violate that right.

But maybe there are other reasons for thinking that Nicholas has a right

13 Thomson, *The Realm of Rights*, 157.

against me that I not press my doorbell. Perhaps Nicholas's having that right is necessary to explain why forcible action may be taken, by Nicholas and others, to prevent me from pressing my doorbell. After all, normally, others may not forcibly act to prevent me from pressing my doorbell. But, again, we need not invest in Nicholas a distinct right against me that I not press my doorbell to explain this. All we need appeal to is the fact that forcibly preventing me from pressing my doorbell is both a necessary and proportionate means of preventing me from violating Nicholas's right against me that I not kill him. Furthermore, suppose that I am whistling as I walk toward the doorbell. If forcibly preventing me from whistling were both a necessary and proportionate means of preventing me from violating Nicholas's right against me that I not kill him—because, say, doing so would distract me long enough to allow Nicholas to detach himself from the battery—then doing so would be permissible. But, note, Nicholas certainly does not have a right against me that I not whistle (nor could MPR deliver the result that he has any such right, for were I to kill him I would not do so *by* whistling). So the supposition is misguided that Nicholas must have a right that I not do whatever preventing me from doing would be both a necessary and proportionate means to preventing me from violating his right that I not kill him. Rather, the explanation is far simpler: *X* may ϕ if ϕ -ing is a necessary and proportionate means to preventing *Y* from violating *X*'s right that *Y* not kill *X*.

Thomson's other reason for thinking MPR is true is that it can explain the truth of:

The Means Principle for Permissibility (MPP): If (i) if *Y* does *a* then *Y* will do β , and (ii) it is permissible for *Y* to do *a*, then it is permissible for *Y* to do β .

Again, this is unpersuasive. Even if MPP is true, MPR might well be false. MPR also seems superfluous to the explanation Thomson offers. If a right is infringed in achieving an end, we need not appeal to some distinct right infringed in implementing the means to it to explain the permissibility facts. We can simply appeal to the right infringed in achieving the end those means are means to to do that.

Furthermore, the postulation that MPR is that which explains the truth of MPP is dubious, for MPP clearly applies even in cases in which the permissibility facts are not a function of rights. Suppose *X* can press a blue button, thereby saving *A* from drowning, or she can press a red button, thereby saving *A*, *B*, and *C* from drowning. In this case, neither *A*, *B*, nor *C* has a right against *X* that *X* rescue him.¹⁴ Nevertheless, MPP clearly applies here: it is morally impermissible for *X* to press the blue button in this case because it is not morally permissible for

14 That there is no right to be rescued is something for which Thomson argues explicitly and persuasively in *The Realm of Rights*, 160–63.

him to only save *A*, and if he presses the blue button he will thereby only save *A*. As no rights are in play and yet MPP applies in this case, what explains the truth of MPP must be something far more general than any principle like MPR, which only concerns rights, specifically.

What's more, double-counting worries may well threaten the very compatibility of MPR and MPP. Suppose bringing about a good, *G*, justifies infringing a right, *R*. *G* might be sufficient to justify infringing *one* right of the stringency of *R*, but insufficient to justify infringing *two* rights of *R*'s stringency. If so, we get the very odd result that if I can infringe *R* directly to bring about *G*, then infringing it would be permissible, but if I could do so only by doing something else, then infringing it would not be permissible. MPR and MPP make uncomfortable bedfellows. MPR does no work that could not be done without it, and it creates unnecessary problems.

I cannot see, nor has Thomson offered, any other reason for thinking that there are the special kinds of rights MPR postulates. There is no explanatory work that needs doing that cannot be done simply and efficiently without them, and were they to exist they would be unlike all other rights in many diverse ways. I conclude both that we have no reason to accept MPR and that parsimony considerations, in addition to the fact that the rejection of MPR offers the prospect of a rights-based solution to the Trolley Problem, give us good reason to reject it.

6. THREE OPTIONS AND THE PRO TANTO OBLIGATION TO MINIMIZE RIGHT INFRINGEMENTS

Before offering my own preferred solution to the Trolley Problem, I want first to briefly revisit Thomson's discussion of *Three Options*. Though it is inessential to her ultimately failed argument in "Turning the Trolley," I think it is illuminating nonetheless.

Three Options supposedly motivates (*) and *Third Principle*. It does not. I agree that only (i) and (iii) are permissible in *Three Options*: the bystander may not turn the trolley onto the one in that case. However, neither (*) nor *Third Principle* is part of the explanation why. Consider:

Shepherd (Three Options): Things are as they are in *Shepherd* except that the shepherd has a third option, (iii): he can jump onto the track in front of the five. If he does, though he will be killed, he will save the five because his body will bring the trolley to an abrupt halt.

All of the shepherd's options—(i) do nothing, (ii') pull the five off the track, and (iii) jump in front of the trolley—are permissible. It is not the case, then, that if

doing a good deed has a cost and a person can pay it himself, then, if he does the good deed, he must pay the cost himself.

So (*) is false. *Third Principle* is, too. Consider:

Three Options (Consent): Things are as they are in *Three Options* except that if the bystander flips the switch to the right (option (ii)), the one on whom he turns the trolley has freely and rationally consented to having it turned on him to save the five.¹⁵

In this case, the bystander may take any of his three options. Sometimes it is permissible to kill one to save five even if one can instead kill oneself to save them.

Thomson might cry foul here. *Third Principle* was only ever offered as a *ceteris paribus* principle, and in *Three Options (Consent)* *ceteris* are not *paribus*. In that case, the one has consented to the trolley being turned onto him and perhaps that kind of complication was meant to be excluded by *Third Principle's ceteris paribus* clause.

Now that may be, but I think it obscures a deeper truth. For though *Third Principle's ceteris paribus* clause may catch *Three Options (Consent)* in its net, that case has something in common with *Three Options* that explains *why* the data are as they are in those cases. In both, the one whom it is permissible to kill has no right against the bystander that he not kill him to save the five. In *Three Options (Consent)* the one has waived that right by consenting to being killed, and in *Three Options* there is no such right because the bystander does not have such a right *against himself*.

When I chop off your finger, I infringe your right against me that I not harm you. What follows from this? Many things, including:

- (1) to be permissible (a) the good that I achieve in harming you must be substantially greater than the loss you suffer, (b) perhaps certain particular causal relations between my harming you and the good thereby achieved must obtain, (c) ...
- (2) prior to harming you I must, other things equal, seek a release from you to harm you, and
- (3) after harming you I must, other things equal, compensate you for your loss.¹⁶

¹⁵ All of the points I go on to make in what follows could equally well be made concerning a version of *Three Options* in which option (ii) involves the bystander turning the trolley onto the one who villainously launched it toward the five in the first place.

¹⁶ An in-depth discussion of these and many other aspects of rights may be found in Thomson's *The Realm of Rights*.

These are (some of) the hallmarks (defeasible, to be sure) of the right not to be harmed. They all apply whether I chop off your finger or kill you by turning a trolley onto you. This is not so when it comes to harming myself, however. To harm myself permissibly the good I bring about in doing so need not be greater than the harm I suffer. Also, I need not seek a release from myself prior to harming myself, nor do I owe myself compensation afterward. In the case of self-harm, all of the hallmarks of a right not to be harmed are absent. That is as it should be; we do not have any such rights against ourselves. What is more, I submit, this is an important part of the special moral authority we each have, and no one else has, over our own bodies.

Instead of *Third Principle*, then, to explain *Three Options* we can instead appeal to

Rights Principle: A must not kill B to save five if (a) doing so would infringe a right of B's against A that A not kill B, and (b) A can instead kill someone who does not have a right against A that A not kill him to save the five.

Rights Principle explains what is going on in *Three Options* in a way that is consistent with the data in *Three Options (Consent)* and is more unifying than *Third Principle*. The bystander may do nothing (option (i)) for one may let five die if every other option would lead to the death of someone else who otherwise would not. He may also turn the trolley onto himself (option (iii)), for one may sacrifice oneself to save five others. And he may not take option (ii) because, in accordance with *Rights Principle*, he has at least one option available to him (option (iii)) whereby he saves the five by killing someone (himself) who does not have a right against him that he not kill him.

Rights Principle is not only more unifying than is *Third Principle*, it is also more explanatorily powerful. Take the following variant of *Three Options (Consent)*:

Three Options (Consent)': Things are as they are in *Three Options* except that the bystander's options are: (i) do nothing; (ii) flip the switch to the right, thereby killing the one who has not consented to having the trolley turned onto him; and (iii) flip the switch to the left, thereby killing someone else who has consented to having the trolley turned onto him.

Rights Principle, unlike *Third Principle*, can explain the moral impermissibility of option (ii) in this case, and it can do so in a manner exactly parallel to the way it does in *Three Options*. Option (i) is permissible for the same reason as it is permissible in *Three Options*. Option (iii) is permissible because, as seems intuitive,

it is permissible to kill one to save five if the one who is killed has no right against the bystander that he not kill him. And given the permissibility of option (iii), the impermissibility of option (ii) follows via *Rights Principle* just as it does in *Three Options*.¹⁷

Rights Principle is itself an instance of the more general *ceteris paribus* principle:

Minimize Right Infringements: A ought not to ϕ in achieving result, R, if (a) doing so would infringe a right of stringency, S, and (b) A can instead achieve R without infringing a right of stringency greater than or equal to S.¹⁸

Minimize Rights Infringements is consonant with the picture of rights as means-blocking barriers I mentioned above and this further recommends it as a component of the full story about the ethics of harming in these cases. *Minimize Rights Infringements* embodies the thought that rights constrain the means by which we may harm people to minimize harm overall not only in virtue of those means' causal properties, but also in virtue of their relational properties. It is an affront to one's moral integrity to be harmed to minimize harm overall via certain causal routes; it is also an affront to be harmed to minimize harm overall when harm overall could be minimized without treading upon one's rights. *Minimize Rights Infringements*, then, is a natural elaboration of the very conception of rights to which First Solution makes appeal.

Three Options is explicable, but not by way of any principles that ground an argument for the impermissibility of the bystander's taking option (ii) in *Bystander*. The impermissibility of the bystander's taking option (ii) in *Three Options* is explained by *Rights Principle*, which is just an instance of *Minimize Right Infringements*. What is more, these principles are consistent with the intuitive verdict

17 Does *Rights Principle* not entail in *Shepherd (Three Options)* that (ii) is impermissible given that (iii) is permissible? No. It would do so only if in pulling the five off the track, the shepherd would thereby infringe the right of the one standing behind them that he not kill him. But it is not at all clear that in pulling the five off the track the shepherd does kill the one. And even if it does count as a killing, such a killing is not the kind of killing we, in general, have rights against others that they not perpetrate against us. If I duck a bullet that kills you but would not have killed you had I not ducked to save my own life, then, whether we say I have killed you or not, I have not infringed any right of yours in ducking. The morality of ducking is discussed in Boorse and Sorensen, "Ducking Harm."

18 Interestingly, this is a principle Thomson not only might endorse, but has endorsed explicitly: "However weak a [right], a large increment of good to be got by infringing it does not make infringing the [right] permissible if one has a non-[right]-infringing way of producing the same increment of good, or even a comparably large increment of good" (Thomson, *The Realm of Rights*, 164).

that the bystander may take option (ii) in *Bystander*, for in that case the bystander has no non-right-infringing way of preventing the greater harm in question.

7. TOWARD A SOLUTION TO THE TROLLEY PROBLEM

Thomson's reversal on her own Trolley Problem is ill motivated. The principles to which she appeals in her argument that option (ii) in *Bystander* is impermissible are false. What is more, the moral data in *Three Options*—data that I do not question—can be accommodated by *Rights Principle* and *Minimize Right Infringements*, principles consistent with the permissibility of the bystander's taking option (ii) in *Bystander*. The Trolley Problem, then, is alive and well and in need of a solution. Thomson's Second Solution fails and her grounds for ditching her First Solution in favor of it are bad ones. I think it is correct to place rights infringements at the heart of the solution, as First Solution does, but First Solution does not employ rights infringements in precisely the right way. Here I will sketch what seems to be a more promising way of incorporating rights infringements into a solution to the Trolley Problem.

Recall First Solution: the bystander may take option (ii) in *Bystander* because in doing so he (1) makes what was threatening five come to threaten only one, and (2) does so not by any means which, in themselves, constitute an infringement of any right of the one's. The problem with First Solution is that there are cases in which the moral data are not as First Solution would have them. Here are two cases in which it is permissible to harm the few in order to save the many that do not involve making what was threatening the many threaten the few:

Trolley (Avalanche): Everything is as it is in *Bystander* except that if the bystander takes option (ii) the trolley will be redirected onto an empty spur of track and collide with a cliff wall thereby causing an avalanche that will crush and kill one person trapped below.

Trolley (Landslide): Everything is as it is in *Bystander* except that the bystander cannot redirect the trolley onto the side spur of track. But he can activate a device (option (ii)) that will launch the five safely onto a bluff high above the tracks. Their landing there, however, would cause a landslide that would crush the one trapped below.¹⁹

In both cases the bystander may cause the lesser harm, but in neither does he make what was threatening the five threaten the one. Rather, both cases involve

19 These are variants of cases presented in Kamm, *Intricate Ethics*.

the creation of a new threat to the one.²⁰ True, First Solution only purports to offer a sufficient condition for harming some to save others, and so these cases are not, strictly speaking, counterexamples to it. They do demonstrate, however, that First Solution is not the complete story of the permissibility of harming some to save others and they can help point the way toward a more complete one.

Here is a case in which redirecting a threat from the many onto the few by means which themselves are not an infringement of anyone's rights is impermissible:

Trolley (Two Effects): Everything is as it is in *Bystander* except that the one is not trapped on the side spur of track but is instead standing alongside it. If the bystander presses a button (option (ii)), his doing so will have two distinct effects: (1) it will cause the trolley to be redirected onto the side spur of track and (2) it will activate a machine that will push the one onto the side spur of track right in front of the oncoming trolley (in such a way that he will be unable to avoid being run over and killed by it).

The bystander may not take option (ii) even though he would be redirecting a threat away from the many onto the few by means that would not infringe anyone's rights—the one has no right that the bystander not press the button nor that he not redirect the trolley. *Trolley (Two Effects)* is thus a counterexample to First Solution.

A slightly different solution is called for. I tentatively propose:

Harm Prevention Principle (HPP): One may cause a right-infringing harm, *h*, that otherwise would not have occurred only if

- (i) the total harm caused is less than would occur were *h* not caused,
- (ii) *h* is caused by some event, *P*, that is the prevention of some greater harm, *H*, and

20 *Trolley (Avalanche)* and *Trolley (Landslide)* also make trouble for the solution to the Trolley Problem offered in Haslett ("Boulders and Trolleys"). That it is permissible to harm others by creating new threats in these versions of the trolley scenario casts doubt upon solutions to the Trolley Problem, like Haslett's, which permit only *shifting* dangers away from the many onto the fewer, and not *originating* dangers to the fewer in order to save the many. That these cases undermine such solutions is significant because Haslett's solution, like the solution I go on to offer, is an attempt to develop and strengthen Thomson's First Solution to the Trolley Problem. Thanks to an anonymous referee for impressing upon me the significance of *Trolley (Avalanche)* and *Trolley (Landslide)* for Haslett's solution to the Trolley Problem.

- (iii) any infringement of the serious rights of those who suffer harm in the course of the causing of h neither causes P nor causes h .²¹

According to HPP, causing a lesser harm can only be justified if that harm is a causal consequence of the prevention of some other greater harm, i.e., the prevention event, and also—and here is where HPP borrows from Thomson's First Solution—any infringements of the serious rights of those who suffer the lesser harm are neither a cause of the prevention event nor a cause of the lesser harm. According to HPP, then, neither the prevention event itself, nor any of the events that either cause it or cause the lesser harm, may be something the one harmed has a serious right against the one causing the harm that he not do. Causing a lesser harm is morally permissible, then, only if the lesser harm lies causally downstream from the prevention of the greater harm, and no serious right infringement of the one who suffers the lesser harm lies causally upstream from either the lesser harm or the prevention of the greater harm.

- 21 HPP is inspired in large part by considering many of the cases Frances Kamm introduces in *Intricate Ethics*, and reflecting upon her discussion of them. Though HPP is similar to Kamm's theory in many ways, the theories are distinct. For Kamm, what is crucial is that the lesser harm not be causally upstream from a "greater good." HPP requires that all lesser harms be causally downstream from the greater harm prevented. (This makes for a significant difference with Kamm's theory, for a lesser harm can fail to be causally upstream from the greater harm prevented without being causally downstream from it.) HPP also does not allow harming in order to bring about goods, however great they may be, that are not the prevention of greater harms. It may be that Kamm's theory (which is too complex to state in detail here) and HPP are fastening onto the very same phenomena. (Whether they are may well depend on whether by the "greater good" Kamm just means what I have identified as the "prevention event," though Kamm's use of the "greater good" does not settle the matter.) I do not, therefore, offer HPP here as a rival to Kamm's theory. However, in structure, it is quite a bit simpler than is hers. HPP, though it does crucially appeal to the notion of a prevention event, does not make recourse to any of the machinery that figures centrally in Kamm's theory, machinery that includes such notions as: "means to the greater good," "non-causal flipside (of the greater good or means to the greater good)," "the structural equivalent of the greater good," and "means which overlap the involvement-of-the-person part of evil.*" Though precise necessary and sufficient conditions for an event being a prevention event are hard to come by, I think it is not so difficult to pick out in various cases which event is the prevention event. By contrast, Kamm's notions, listed above, are not so easy to identify across cases. Also, like Thomson's theory and unlike Kamm's, rights figure centrally in HPP. This is also a merit of HPP, for the infringement of rights, as Thomson highlighted when she presented her First Solution, does seem to play an important role in the correct explanation of which harmings are and are not permissible. Last, HPP delivers different, and I believe correct, verdicts from those delivered by Kamm's theory in a number of Kamm's own cases: Lazy Susan (With Man Trapped Under It), Vibrations, Component Case II, Worse Than Dead, and Sending Back. (See Kamm, *Intricate Ethics*.) For these reasons, I prefer HPP to Kamm's theory. (HPP is also meant to be supplemented by *Minimize Right Infringements*, of course.)

HPP delivers the intuitively correct verdicts in the cases discussed so far. First, here are the prevention events in each of the cases in which it is permissible to harm the one:

CASES	PREVENTION EVENT
<i>Driver, Bystander, and Trolley (Avalanche)</i>	→ the turning of the trolley
<i>Shepherd</i>	→ the removal of the five from the track
<i>Trolley (Landslide)</i>	→ the launching of the five off the track

In each of these cases all lesser harms caused are caused by the prevention event and no serious right infringements cause either the prevention event or a lesser harm. Thus, HPP declares harming permissible in each of these cases. Next, here are the prevention events in each of the cases in which it is impermissible to harm the one:

CASES	PREVENTION EVENT
<i>Big Man</i>	→ the trolley colliding with the big man
<i>Trolley (Two Effects)</i>	→ the turning of the trolley

In each of these cases there is an infringement of the serious rights of the one harmed that either causes the prevention event or causes the lesser harm. Thus, HPP declares harming impermissible in each of these cases.

Absolutely crucial to the application of HPP is there being some event in the course of the permissible causing of a lesser harm that is the prevention of a greater harm. In each of the cases above I identified which event is the prevention of the greater harm in that case and noted that the moral data in the case follow from HPP and the event so identified. Though it seems clear in many cases what the prevention event is, it is no easy task to give necessary and sufficient conditions for an event's being the prevention of some harm. That said, however, it is not that there is nothing that can be said.

First, it is certainly not sufficient for an event's being the prevention event that it be such that, had it not occurred, the greater harm would have occurred. That counterfactual is true for many events that are not the prevention event. For instance, the bystander's pressing the button in *Trolley (Two Effects)* (or the thin man's pushing the big man in *Big Man*) is an event that is such that had it

not occurred the greater harm would have occurred, and yet it is not the prevention event, the turning of the trolley is (the big man's halting the trolley is). But, though not sufficient, the holding of that counterfactual does seem to be a necessary condition for an event's being the prevention event. How could an event be that in virtue of which a harm is prevented if it was not such that, had it not occurred, the harm would have occurred? Another thing that seems necessary is that the event involve either the process by which the harm would occur or the victims of that harm were the harm not prevented. Whenever a harm occurs there is some identifiable process by which the harm occurs and some victims who suffer it. For an event to be the prevention event, it seems, it must be an event involving either that process or its potential victims.²² For instance, the reason why the bystander's pressing the button in *Trolley (Two Effects)* (or the thin man's pushing the big man in *Big Man*) is not the prevention event is precisely because it is not an event involving the process by which the harm would occur were it not prevented; in *Trolley (Two Effects)* (or in *Big Man*), were the harm to the five not prevented, the harm would be caused by the trolley barreling into the five, and so the process by which that harm would occur if it were not prevented is the process of the trolley's barreling toward them.

The distinction I am drawing here—that between an event by which one prevents a harm and the prevention event—can be seen even in cases in which agency plays no role:

Trolley (Breeze): An out-of-control trolley is barreling toward five track workers who are trapped on the track ahead of it. If nothing stops the trolley, the five will be run over and killed. A gentle breeze dislodges a pebble resting along the cliff face. The pebble bounces down the side of the cliff and strikes a boulder stuck in the cliffside, causing it to roll down the cliff. The boulder collides with the trolley and knocks it off course. The five are saved.

In this case, we can say that the gentle breeze prevented the harm to the five by dislodging the pebble (and that the pebble prevented it by dislodging the boulder), but the breeze's dislodging of the pebble is not itself (nor is the dislodging of the boulder) the prevention event. As in many of the other trolley cases, the

22 It is a further merit of HPP, in contrast with other solutions to the Trolley Problem (including Kamm's), that it focuses on the prevention event. Once it is recognized that whenever a harm is prevented there is a process whereby the harm prevented would have occurred had it not been prevented it naturally falls out, as a matter of course, that all permissible harmings fall either into one class—interferences with an independently identifiable harm-causing process—or another—interferences with the potential victims of the harm-causing process.

prevention event is an event involving the process by which the harm would have been caused were it not prevented, viz., the event of the trolley's being knocked off course by the boulder.

A prevention event can sometimes be an event not involving the harm-causing process; it can also sometimes be an event involving the potential victims of the harm whereby they are rendered no longer susceptible to the harm that the harmful process would cause were it not prevented. So, in *Trolley (Landslide)* the prevention event is the event of the victims being launched out of the way of the oncoming trolley. Here is another example:

Trolley (Tractor): Everything is as it is in *Bystander* except the bystander cannot redirect the trolley away from the five. Instead, he can (option (ii)) launch a tractor that will gently push the five out of the way of the oncoming trolley. Unfortunately, an innocent person is trapped between the tractor and the five, and so if the bystander launches the tractor, it will run over and crush the one en route to pushing the five out of the way of the trolley.

In this case and in *Trolley (Landslide)* if the bystander takes option (ii), the prevention event is the event involving the potential victims whereby they are rendered insusceptible to the harm. In *Trolley (Landslide)* the prevention event is their being launched to safety, and in *Trolley (Tractor)* it is their being pushed out of the way by the tractor. But whereas the bystander's taking option (ii) in *Trolley (Landslide)* is permissible, his taking option (ii) in *Trolley (Tractor)* is not, and HPP explains why: in the former, the lesser harm the bystander causes to the one is caused by the prevention event, whereas in the latter, it is not (rather, it is caused by an event that is a cause of the prevention event).²³

- 23 It is not clear whether, according to HPP, turning the trolley is permissible in Thomson's Loop Variant of *Bystander* ("The Trolley Problem"). In this variant (call it *Loop*), the side track, on which the one is trapped and onto which the bystander can redirect the trolley, loops around and reconnects with the main track headed toward the five. If the bystander redirects the trolley onto the one, the one's being killed will halt the trolley, thereby preventing it from looping back around and running over and killing the five. Though I am inclined to think that it is permissible to redirect in *Loop*, I grant that one might reasonably disagree. In fact there are two simple and natural alternative additions to the partial characterization of what it is for an event to be the harm prevention event I have offered, each of which delivers one of the two possible verdicts about redirecting the trolley in *Loop*:

Addition #1: An event is the harm-prevention event only if it is the *earliest* event involving either the harm-causing process or the potential victims of the harm such that had it not happened the harm would have occurred.

Addition #2: An event is the harm-prevention event only if it is the *latest* event in-

What more can be said about what it is for an event to be the event in virtue of which a harm is prevented? It is not clear. The notion of a prevention event is no more precise than that either of a threat or of the means by which a threat is redirected, and so HPP has no advantage over Thomson's First Solution in terms of precision. Though in many cases it is intuitively clear which event is the prevention event, in many other cases this is not so clear. With respect to these latter cases, so long as it is also unclear whether it is permissible for the agent in question to proceed, they constitute no grounds for rejecting HPP.

I have argued that HPP can explain the moral data in cases in which agents do harm in order to prevent greater harm to others. But even if HPP is correct, one might wonder: *why* is it correct? One reply—the easy reply—is that it is correct because it best accounts for the data. To an extent this is a good reply; what better test of an explanation is there, after all, than whether it best accounts for the data? On the other hand, this reply might leave one cold. One might want something more unifying, that which “lies behind” HPP.

I think that which lies behind HPP is a thought that might plausibly be thought to be at the bedrock of non-consequentialist moral theory. The thought is that which is encapsulated in the foremost of the prescriptions of the Hippocratic oath, viz., “do no harm.” Now HPP does indeed allow harming in certain cases, but it does so only when the harms it allows are ones that are the causal upshots of the prevention of yet greater harms. And this is because, in a way, when the lesser harm is caused by the prevention of the greater harm, what one does in jointly preventing the greater harm and causing the lesser harm is to *transform*

volving either the harm-causing process or the potential victims of the harm such that had it not happened the harm would have occurred.

With Addition #1, redirecting the trolley in *Loop* is morally permissible; the earliest event involving the harm-causing process such that had it not happened the harm would have occurred is the turning of the trolley, and if that is the harm-prevention event, then HPP declares turning the trolley permissible. With Addition #2, redirecting the trolley in *Loop* is morally impermissible; the latest event involving the harm-causing process such that had it not happened the harm would have occurred is the trolley's hitting the one, and if that is the harm-prevention event, then HPP declares turning the trolley morally impermissible. That the framework provided by HPP can explain why it is that the moral data in *Loop* are much disputed may well be another mark in its favor over other rival solutions to the Trolley Problem—each of Addition #1 and Addition #2 is a natural and plausible filling in of the account of what it is for an event to be the prevention event, and so, perhaps, what is at the root of the dispute between proponents and opponents of the permissibility of turning the trolley in *Loop* is their respective implicit acceptance of one or the other of Additions #1 and #2. And because Addition #1 and Addition #2 are both easily incorporatable into an account of what it is to be the prevention event, HPP is more flexible than other solutions, like Kamm's, for instance, which take the permissibility of turning the trolley in *Loop* as a datum.

the greater harm into the lesser harm. On this way of thinking, when one causes the lesser harm, one is not so much introducing a new harm into the world as transforming a greater harm into a lesser one. And if this is right, then HPP can be seen as a slight modification of the very intuitive thought that we are morally required to avoid harming others. The modification that this interpretation of HPP suggests is that what morality requires is not, strictly speaking, that we not harm others, but that we avoid introducing *new harms* into the world, where a new harm is one that has not been transformed from some other harm. This, if right, can explain the particular requirement that the lesser harm be *caused by* the prevention of the greater harm. This is because, you might think, a greater harm is not transformed *into* a lesser one if the prevention of the greater harm is not causally prior to the lesser harm into which it is transformed. For if the lesser harm does not come out of, or emerge from, the prevention of the greater harm, then its existence is antecedent to, and thus, in a sense, independent of, the prevention of the greater harm. And if the lesser harm has an existence antecedent to, and independent of, the prevention of the greater harm, then the greater harm that has been prevented has not been *transformed into* the lesser one. Another way of putting this is to say that morality requires that we not introduce any harms into the world that are not already “paid for” by the overall reduction of harm suffered in the world. (Here, again, the “already” is crucial; it is what accounts for the requirement that whatever harms one causes lie causally downstream from the prevention of a greater harm.)

But the thought that the transformation of a greater harm into a lesser harm can be seen as not really bringing another new harm into the world is only part of that which lies behind HPP. For not only does HPP require that the harm prevented be greater than the harm caused and that the prevention of the greater harm be that which causes the harm caused, but it also requires that any right infringement that occurs in the transformation of a greater harm into a lesser harm not be the cause either of the prevention of the greater harm or of the lesser harm. Here the thought connects up once again with the conception of rights as means-blocking barriers. If in minimizing harm overall one is transforming a greater harm into a lesser harm, then rights will act as barriers to that transformation. And insofar as the transformation of the greater harm into the lesser harm is constituted jointly by the prevention of the greater harm and the occurrence of the lesser harm, the means to the transformation of the greater harm into the lesser harm are the causes of those two events. So rights, being means-blocking barriers to the transformation of greater harms into lesser ones, are moral barriers to causing the two events that constitute the transformation of the greater harm into the lesser harm. And so, for a transformation of a greater

harm into a lesser harm to be permissible, none of the causes of the two events that constitute it, i.e., the prevention event and the lesser harm, may be a serious right infringement of those suffering harm.

Now all of this is, to be sure, sketchy and somewhat inchoate. I have appealed to the fact that a greater harm causing a lesser harm is not "the bringing into the world of a new harm," but merely "the transformation of a greater harm into a lesser one." And I have also claimed that rights being means-blocking barriers to the transformation of greater harms into lesser harms requires that the means to that transformation, i.e., the causes of the two events that constitute the transformation (the prevention event and the lesser harm), not involve any serious rights infringements of those suffering harm. I admit that all of this is, at best, metaphorical. But in trying to offer the thoughts that lie behind whatever fundamental principles we think constitute the core of the morality of harming, we really are striking bedrock. When we reach this depth of explanation, it is hard to provide something more concrete than HPP, and so if we do try to dig deeper, we are bound to drift somewhat toward metaphor.

HPP, then, is my tentative proposal for the principle governing permissible right-infringing harms. It (or some suitably modified version of it) can, I am optimistic, account for the moral data in these and other cases in which harm is done to prevent yet other harms. It also seems to capture an intuitive thought that we are allowed to cause harm only when our doing so transforms a greater harm into a lesser one without steamrolling others' inviolability, as embodied in the rights they have, on the way to doing it.

8. CONCLUSION

Thomson's twist on Foot's original Trolley Problem introduced a particularly difficult puzzle for non-consequentialist moral theory. Though Thomson has of late come to think she was wrong to reject Foot's solution, I have argued that she should not have changed her tune. I have explained where Thomson's argument goes wrong and how the data in the cases she employs to motivate her dismissal of the Trolley Problem as a nonproblem can be accounted for without licensing that conclusion. I have also gestured toward a solution to the problem that seems more promising.²⁴

*University of Massachusetts
pgraham@philos.umass.edu*

²⁴ I would like to thank Michael Otsuka and two anonymous referees for comments on earlier drafts of this paper.

REFERENCES

- Boorse, Christopher, and Roy A. Sorensen. "Ducking Harm." *Journal of Philosophy* 85, no. 3 (March 1988): 115–34.
- Dworkin, Ronald. "Rights as Trumps." In *Theories of Rights*, edited by Jeremy Waldron, 153–67. Oxford: Oxford University Press, 1984.
- Foot, Philippa. "The Problem of Abortion and the Doctrine of Double Effect." *Oxford Review* 5 (1967): 5–15.
- Haslett, D. W. "Boulders and Trolleys." *Utilitas* 23, no. 3 (September 2011): 268–87.
- Kamm, Francis M. *Intricate Ethics*. Oxford: Oxford University Press, 2007.
- . *Morality, Mortality*. Vol. 2, *Rights, Duties, and Status*. New York: Oxford University Press, 1996.
- Smart, J. J. C., and Bernard Williams. *Utilitarianism: For and Against*. Cambridge: Cambridge University Press, 1973.
- Thomson, Judith Jarvis. "Killing, Letting Die, and the Trolley Problem." *The Monist* 59, no. 2 (April 1976): 204–17.
- . *The Realm of Rights*. Cambridge, MA: Harvard University Press, 1990.
- . "The Trolley Problem." *Yale Law Journal* 94, no. 6 (May 1985): 1395–1415.
- . "Turning the Trolley." *Philosophy and Public Affairs* 36, no. 4 (Fall 2008): 359–74.