

## **TROLLEYS AND PERMISSIBLE HARM**

**Thomas Hurka, University of Toronto**

I'd like to start by quoting something that appeared in *The Globe and Mail* a few years ago. The newspaper had run a book review that mentioned the trolley problem, and this prompted the following letter to the editor:

The ethical dilemmas involving a runaway trolley illustrate the uninformed situations that cause people's eyes to glaze over in philosophy class. Trolleys and trains are unlikely to run away because they're equipped with a "dead man's pedal" that applies the brakes if the driver is incapacitated.

The potential rescuer would not have the choice of "throwing the switch" because track switches are locked to prevent vandalism. And the rescuer's response would depend on the speed of the trolley. If the speed were less than 15 kilometers an hour, the rescuer could jump onto the trolley, sound the bell and save all five lives. If the speed were less than 30 km/h, then the rescuer (with a switch lock key) could throw the switch and kill only the one person on the branch line.

If the trolley were moving faster than 30 km/h, throwing the switch would cause it to derail, which would injure or kill the passengers but save the workers on the tracks. So the better choice

is to allow the occupied trolley to run through on the main track  
and, regrettably, kill, the five workers.

*– Derek Wilson, former CN Rail transportation  
engineer and project manager, Port Moody, B.C.*

At the risk of annoying any retired rail engineers who may be reading, I'll take the opportunity afforded me by Frances Kamm's stimulating Tanner Lectures to say something philosophical about trolleys.

Kamm's widely recognized brilliance as a moral philosopher has two sides. She's an incisive critic of other moral views, pointing out implications their proponents didn't realize they have and challenging them with arresting counterexamples. But she's also a creative philosopher who's proposed many novel moral principles and ideas. Her critical side is to the fore in her first Tanner Lecture and her constructive side in her second. I'll focus on a positive claim about the Trolley Problem from her second lecture; this will involve my trying to play the role of critic.

The Trolley Problem asks why it's impermissible to save five people by throwing a fat man in front of a trolley but permissible to save them by diverting the trolley to another track where it will likewise kill one person. This is a problem because the natural explanation of why it's wrong to throw the fat man – that then you actively do something that causes his death rather than merely allow him to die – implies that it would also be wrong to divert the trolley, which isn't what most of us intuitively believe. And while the doctrine of double effect seems to yield the right results here, Kamm thinks it has counterintuitive implications in other cases and should therefore be rejected. Hence her proposal, both in earlier writings and in the first part of her second lecture, of a novel Principle of Permissible Harm that she says avoids these pitfalls.

Like the view that distinguishes doing from allowing, Kamm's Principle looks at the causal process that produces the death of the one rather than at any facts about your mental state. But, unlike that view, it doesn't consider how *you* relate to his death. Instead, it focuses on a causal distinction that arises later in the causal process, after you act. That's one reason for calling it a "downstream" principle: its vital distinction occurs after or downstream from you.

What is this distinction? Both when you throw the fat man and when you divert the trolley your act produces an evil, namely one person's death, but it also produces a greater good, which Kamm describes as "the five being saved." I think this is an ambiguous description, and will return to this issue later; for now I'll accept the description. What her Principle of Permissible Harm says is that it matters how the lesser evil is caused. If this evil results from a causal means to the greater good, then the act producing the greater good is impermissible. That's why throwing the fat man is wrong. He'll be killed by his body's colliding with the trolley, which is a causal means to stopping the trolley and thereby saving the five. But an act that produces the same evil may not be impermissible if the evil results from the greater good itself or from something that's not a causal means to it but instead has the greater good as what Kamm calls its "non-causal flip side," so the two are in effect the same thing. That's why, according to her, diverting the trolley is permissible. Here what saves the five is the trolley's moving away down the other track, and that's not a causal means to their being saved but the same thing under a different description; as she puts it, "the five being saved is simply the trolley moving away" (Lecture II, p. 6). In both the fat man and diversion cases you actively do something, and your act leads, by a series of intermediate steps, to a good and also to an evil, namely the death of the one. In the first case the evil results from one of the intermediate steps, which is why throwing the fat

man is wrong. But it in the second case it results from the good itself or from something effectively equivalent to it, which makes diverting the trolley permissible.

As so applied the Principle of Permissible Harm seems to yield the right result in the trolley cases. But it has what I think are highly counterintuitive implications in other cases and is therefore open to the same kind of objection Kamm makes to the doctrine of double effect. A case of this type first occurred to me while I was thinking about the ethics of war, so let me start with it.

There's a munitions factory that's a legitimate military target and that we want to bomb, but that has civilians living near it who'll be killed by our attack. Most moral views say that if the number of civilians who'll be killed is not disproportionate to the factory's military importance, the bombing is permitted. But Kamm's Principle says that whether that's so depends on how the civilians will be killed. If they'll be killed by a flying piece of the bomb, the bombing is morally impermissible. That's because the good our act can achieve is the destruction of the factory, the bomb's exploding is a causal means to that, and causing even a lesser evil by a means to a greater good is forbidden. But if the civilians will be killed by a flying piece of factory, the bombing is permitted. The factory's exploding isn't a causal means to the destruction of the factory; it's effectively the same thing, or has the destruction of the factory as its non-causal flip side. So the principle that allows the diverting of the trolley also allows bombings in which civilians are killed by flying pieces of factory but doesn't allow bombings in which the same civilians are killed by flying pieces of bomb.

I find this implication very hard to accept. Whatever the morality of collateral harm in war ultimately says, I don't think anything in it can turn on whether civilians are killed by one

kind of flying object rather than another. Nor do I think the world's militaries have a strong duty to develop weapons that will kill civilians only indirectly, by what they make explode, rather than by their own explosive force. Even if Kamm's Principle's yields attractive results in the trolley cases, it seems extremely implausible here.

I presented this case to her at a workshop on the ethics of war some years ago, and in reply she gave me a winning smile and said, "They all laughed at Christopher Columbus." She was acknowledging that her Principle has the implication I pointed out and saying that, however surprising the implication may be, we should accept it. And she's continued to acknowledge that it follows from her Principle. In her recent *Ethics for Enemies* book she says that, if we applied the Principle in war, "it might be permissible to drop bombs on a military factory if the side-effect deaths of civilians were caused by the destruction of the factory itself but not if they were caused by the bombs."<sup>1</sup>

But accepting the implication would be a strange thing for her of all people to do. Her primary method in moral theory has been what she calls the "method of cases." It establishes moral principles by appealing to our intuitive judgements, but the primary intuitions concern particular examples. We don't first decide that certain principles are correct and then settle cases by applying the principles to them. We first consult our intuitions about cases and accept principles only if they match those intuitions.

She'll therefore sometimes reject an attractive-sounding principle because it conflicts with just one or a few particular judgements. In an earlier paper Judith Jarvis Thomson proposed that it's less morally objectionable to redirect an existing threat, as you do when you divert the trolley, than to create a new threat, as when you throw the fat man.<sup>2</sup> Kamm has rejected this

proposal because it yields what she thinks is the wrong result in the Lazy Susan case described in her second lecture, where rotating a lazy susan with five people on it to remove them from the path of a trolley causes a rockslide that kills another person. Even though the rockslide is a new threat, she thinks rotating the lazy susan is permitted.

On a scale of Ingenuity in Philosophical Examples, I'd say this Lazy Susan case comes pretty high up – I certainly couldn't have invented it myself. But precisely for that reason, I find it less than completely decisive. Even if, once I've understood how it works, I share Kamm's intuition that rotating the lazy susan is permitted, I'm not very confident about that intuition just because the example is so far from reality.

But my factory bombing case comes much lower on the scale. In recent years there have been thousands of bombings of military targets in which thousands of civilians have been killed, some directly by the bomb and some indirectly by its effects, for example by the collapse of a targeted building. Yet no one has thought that in these real-life cases Kamm's distinction makes a moral difference. In *Ethics for Enemies* she notes that her distinction isn't recognized in standard just war theory, but I don't think that's only because the theory's framers didn't think of it. I'm pretty sure that if they had thought of it, they would have given it no weight and would have been very confident in doing so. However much force the Lazy Susan case has against Thomson's proposal about redirecting threats, I think the factory bombing case has more force against Kamm's Principle of Permissible Harm.

It may be relevant, however, that the case occurs in war, and this suggests a possible response for her. In *Ethics for Enemies* she suggests that war may be a special context that alters the morality of harming in certain ways, so perhaps a Principle that yields the right results in

trolley cases isn't properly applicable in war, in which case a counterexample set in war doesn't tell against its use elsewhere. And her book sets her Principle aside for just this reason when discussing issues about war.<sup>3</sup>

I don't myself believe that war is a special context. Like others such as Jeff McMahan, I think the morality of war is just an extension of the everyday morality of self- and other-defence. And even if war is a special context, we'd need a specific argument why one of its effects is to make a Principle of Permissible Harm that's otherwise applicable no longer so, and it's hard to see how that argument could go. Finally, war can presumably be a special context only for those involved in it, as soldiers or at least as citizens of a warring nation. But imagine that our enemy has located its munitions factory on its border with a neutral country, so the civilians who'll be killed are neutrals. Our duties to them surely haven't changed, yet I still think it makes no difference whether they're killed by a flying piece of bomb or a flying piece of factory.<sup>4</sup>

Moreover, parallel cases can be constructed outside war and even involving a trolley. Imagine that an out-of-control trolley is hurtling toward me and four other people and, if we do nothing, will kill us. We can't divert the trolley but have a bomb we can throw to stop it. There's no one on the trolley but there's a bystander next to the track who'll be killed if we throw the bomb. May we do so?

People may disagree about this case. I think Thomson will say we may not throw the bomb; we're not permitted to kill the bystander but must instead let the trolley kill us. My own view is that we may throw the bomb, and not just because there are five of us. Even if I were alone on the track I'd be permitted to throw the bomb, because in circumstances like these each person is permitted to care somewhat more about his own good. So I think we may act.

But Kamm's Principle implies that whether or not we may throw the bomb depends on how the bystander will be killed. If she'll be killed by a piece of bomb, we're not permitted to save our lives because then her death will result from a causal means to the stopping of the trolley. But if she'll be killed by a piece of trolley, we are permitted, because the trolley's exploding just *is* its stopping or has its stopping as its non-causal flip side. Kamm implicitly acknowledged this in her first lecture. The case I've just described is close to her Bomb Trolley Case, and about it she said, "it seems to me impermissible to set off a bomb that will stop the trolley from hitting the five when a piece of the bomb will kill a bystander as a side effect" (Lecture I, p. 5; also pp. 6-7). Why the reference here to a piece of the bomb killing the bystander? Why not just say you may not set off the bomb? I think Kamm was recognizing that, given her Principle, using a bomb to stop the trolley isn't wrong if the bystander will be killed by a flying piece of trolley.

So her Bomb Trolley Case is exactly analogous to my factory case, and my intuition about it is exactly the same. Just as it makes no difference in the wartime case whether the civilians will be killed by a piece of bomb or a piece of factory, so it makes no difference in the trolley case whether the bystander will be killed by a piece of bomb or a piece of trolley. I don't say this just because I think throwing the bomb is permissible. I suspect that Thomson, if she thinks throwing the bomb is impermissible, will agree that it doesn't matter whether the bystander is killed by one flying object rather than another. And if that's right, then it's not true that Kamm's Principle has counterintuitive implications only in wartime, when it might be argued that it doesn't apply; it also has them in cases involving trolleys. Though it may fit our intuitions about the two cases in the original Trolley Problem, it doesn't fit our intuitions in other

cases and therefore fails in the same way as the doctrine of double effect.

So far I've followed Kamm's "method of cases" and criticized her Principle of Permissible Harm for failing to match some particular intuitions. But I don't think that's the only method we should use in moral theory. Like Shelly Kagan, I think our moral principles should not only match our particular judgements but also be intuitively appealing in themselves; in fact they must be appealing in themselves if they're to *explain* the particular judgements. Here I see a further difficulty for Kamm's Principle as she applies it: the downstream causal distinction it draws doesn't seem in itself morally significant.

The traditional deontological principles concern your relation to an evil: they say it's more objectionable to cause than to allow an evil or to intend than merely to foresee it. And the relevant evils are ones that matter in themselves, so they involve either the existence of an intrinsic evil, such as a victim's pain when you torture him, or the destruction of an intrinsic good, as when you kill him. Kamm's Principle likewise concerns the causing of evils that matter in themselves, such as the deaths of the one in the trolley cases. And I think her Principle will be most plausible if the goods it permits the evils to follow from likewise matter in themselves, for example by being states that are intrinsically good. If being caused by a good is so important, shouldn't the good in question itself be one that's important?

This condition is satisfied in a case from later in Kamm's second lecture, where five people who cure themselves from a fatal condition then breathe normally and their expanding chests move some fatal germs in the atmosphere that kill another person (Lecture II, p. 22). Here the death of the one results from the five's being alive, which matters morally in itself. But the condition isn't satisfied in other cases, most obviously in my factory bombing case.

Here Kamm's Principle takes the relevant greater good to be the factory's being destroyed, but that's not something of intrinsic moral significance. The world isn't any better just because a factory no longer exists. Instead, the factory's destruction matters only as a means, more specifically as a means to the ultimate goods in the war's just cause, such as ending an unjust aggression or preventing a genocide. It's because and only because it will further a good like this that the destruction is worth pursuing. In fact, we can describe a sequence of means leading to this ultimate good, including the plane's taking off, the pilot's pressing the button that releases the bomb, the bomb's exploding, the factory's being destroyed, and then whatever further means to our end the factory's destruction allows. Perhaps there are a hundred means in this sequence, with the bomb's exploding number 37 and the factory's being destroyed number 38. What Kamm's Principle does, when applied to this case, is to say that an act that kills civilians in pursuit of the ultimate good is permissible if the deaths result from means 38 but not if they result from means 37 or earlier. Why is that reasonable? Why say causation by one means to an end is permissible but causation by another means to the very same end is not? And why make the cut exactly between 37 and 38? Why not between 36 and 37, so the bombing is forbidden if the civilians' deaths are caused directly by the pilot's pushing the button but not if they're caused by the "good" of his bomb's exploding? Or why not make the cut after 38? It seems that selecting any means in a sequence of hundred and labeling it a morally significant "good" is arbitrary and unmotivated.

It may be replied that this again shows why my bombing case isn't relevant to Kamm's Principle. Just because the case doesn't involve a good of the right, intrinsically mattering kind, it's not one to which the Principle applies. But this reply only makes the Principle look worse. If

the factory's destruction is just a causal means to the truly relevant good, then bombing the factory is wrong by Kamm's lights both when the civilians are killed by a piece of bomb and when they're killed by a piece of factory. In fact, killing civilians is then permitted only when it results from the final achievement of the war's just cause, which means it's never permitted while the war is ongoing. I take it that's not an implication Kamm wants her Principle to have.

The same difficulty arises in the trolley cases, which brings me back to the question of what the greater good in those cases is. Kamm, recall, describes the good in the fat man and diversion cases as "the five being saved." But what exactly does that mean, and is it something that matters morally in itself or only as a means?

Different readings of the phrase are possible, but at one point Kamm equates the five being saved with "the five being alive" (Lecture II, p. 6; also pp. 8, 23, 24), and if this was the greater good it would be something that matters in itself. But we need to ask *when* the five would be alive, or, if your intervention causes them to be alive, *when* it causes them to be alive. To sharpen the issue, imagine that the trolley will reach the divide in the track at 8:00, and that if you don't divert it, will then travel for ten minutes down its original track until it reaches and kills the five at 8:10. Given these facts, the greater good you achieve by diverting the trolley can't be their being alive at 8:01, 8:04, or 8:07, since even if you don't divert the trolley they'll be alive at those times. It must be their being alive at 8:10 and after. But I don't see how their being alive at 8:10 can just "be" the trolley's moving away, or be something the trolley's moving away has as its non-causal flipside. Two outcomes related in that intimate way must surely be contemporaneous, or happen at the same time. But the trolley's moving away happens at 8:00 and the five's being alive at 8:10 happens at 8:10, and nothing that happens at 8:00 can be

identical to or constituted by something that happens at 8:10. The trolley's moving away at 8:00 must therefore be a causal means to the five's being alive at 8:10, and then, given this reading of "the five being saved," Kamm's Principle doesn't permit you to divert the trolley. What causes the death of the one, namely the trolley's moving away, isn't the greater good or something constitutively related to it but a causal means to it. And the Principle forbids acts in which the lesser evil results from a causal means.

A different reading of "the five being saved" equates it with the five's being saved from the threat of the trolley, or being out of danger from the trolley, and Kamm sometimes talks in this way too (Lecture II, pp. 8, 19). Now this good *is* plausibly seen as effectively equivalent to the trolley's moving away, or as something the trolley's moving away has as its non-causal flipside; certainly the five's being saved from the threat of the trolley happens at the same time the trolley moves away, namely at 8:00. But on this reading the greater good isn't something that matters morally in itself; on the contrary, it matters only as a means. To see this, imagine that, no matter what you do with the trolley, a rockslide will come and kill the five at 8:10. Here saving the five from the threat of the trolley is pointless and diverting the trolley would be wrong, because it would needlessly kill the one. So that the five are saved from the trolley matters only if it causes them to be alive at 8:10, and the same is true of their being free from any other threat or indeed from threats in general: all of these matter only as means to something like their being alive. That means the situation given this reading is just like that with means 37 and 38 in the factory case. Imagine that to divert the trolley you have to pull a long handle, one so long that pulling it will knock a thin man off a bridge and kill him. Here Kamm would say that diverting the trolley is wrong because the thin man's death will result from a handle-pulling that's just a

means. But if the five's being saved from the threat of the trolley is likewise just a means to what really matters, why think a death resulting from it or from something effectively equivalent to it is any different? What justifies giving different means to the same end such different moral status?

There are, then, two readings of Kamm's phrase for the greater good, "the five being saved." On one reading this good matters morally in itself, but the trolley's moving away is only a causal means to it and her Principle forbids diverting the trolley. On the other reading the trolley's moving away isn't a causal means, but the so-called greater good doesn't matter in itself and is itself only a means to what does. On this reading the Principle can only permit diverting the trolley by drawing an arbitrary distinction between different means to the same end. On neither reading is it true both that the Principle of Permissible Harm yields the results Kamm wants in the Trolley Problem *and* that it's intuitively appealing in itself.

I've criticized Kamm's Principle both because it has counterintuitive implications in the factory and trolley bombing cases and because it's not intrinsically appealing. These two criticisms are connected. The problem in the bombing cases is that the Principle attaches moral significance to something that doesn't have it, namely the difference between one flying object and another. The reason why the Principle isn't appealing is that it makes an arbitrary distinction between items in a sequence of means. And that's precisely what it does in the bombing cases: make an arbitrary distinction between flying objects that are means to the same good end.

But I don't want only to criticize Kamm, and will conclude with some more positive comments on her Principle and the ideas behind it.

My second criticism concerned cases where what Kamm calls the "greater good" is just a

means to what really matters, and it therefore doesn't apply to applications of the Principle where that isn't the case. I already mentioned one of these: where five people cure themselves of a lethal condition but then their breathing moves some fatal germs that kill another person. In this case the death of the one results from something, the five's being alive, that does matter morally in itself. Moreover, the claim that in this case the five's curing themselves is permissible even though it will kill the one seems to me right, and Kamm's Principle may be the explanation why. I have nothing against the Principle's use in these cases.

In addition, there are cases where an analogous though not identical principle, one that draws a similar downstream causal distinction, does seem morally relevant. Let me describe two of them, from the morality of war.

These cases involve the just war condition of proportionality, which says the resort to war is permitted only if the relevant goods the war will secure are proportionate to, or sufficiently large compared to, the evils it will cause. In this assessment the relevant goods don't in my view include all the goods that the war will produce. They include those in the war's just causes plus some other closely related ones, such as deterrence of would-be aggressors. But they don't include the pleasure our soldiers may get from real action, and they also don't include some economic goods. Imagine that our and indeed the whole global economy is in a recession and that our fighting a war will lift it out of the recession, as World War II ended the Depression of the 1930s. I don't think this economic benefit is relevant to justifying the war; an otherwise disproportionate conflict can't become proportionate because it will boost GDP.<sup>5</sup>

But now imagine that in 1990 Saddam Hussein had occupied not only Kuwait but also the Saudi oilfields and then drastically reduced both countries' oil production, driving up world oil

prices and hurting the economies of African countries. Here the fact that a war against Saddam would remove that harm to African countries, or give them the economic benefit of cheaper oil, does seem to me a relevant good. How can that be if ending our recession wasn't relevant?

Here's my albeit tentative suggestion. The way war lifts an economy out of recession is by calling for more industrial and especially military production, where that additional production boosts GDP; that's certainly how World War II ended the Depression. But industrial production is only a means to the war's just cause rather than part of it, and that's why this economic benefit doesn't count: it's caused by an intermediate step. But in the Saddam case the reduction of oil prices follows from our ending his occupations of Kuwait and Saudi Arabia, which *is* the war's just cause. Since here the economic benefit follows from something good as an end, it is relevant.

Here's another pair of examples. To fight the 1991 Gulf War the US put together a large coalition of nations, including Arab ones and, albeit more informally, Israel, and the resulting contacts between the Arab nations and Israel contributed to an attempt to settle the Israeli-Palestinian conflict through the Oslo Accords. That attempt ultimately failed, but imagine that it had succeeded and led to Israeli-Palestinian peace. Even so, I don't think the good of that peace would have contributed to the proportionality of the 1991 War, and my reason is that what promoted the peace was the formation of the coalition, which was a causal means to the war's just cause rather than any part of it. But now consider the 2003 Iraq War, one effect of which was to stop Iraq's payments to the families of Palestinian suicide bombers. And imagine that the result of this was a cessation of suicide bombing and then Israeli-Palestinian peace. Here I think the good of the peace would have been a relevant benefit, because it would have followed from

the ending of Iraq's support of terrorism, which is a legitimate goal of war.<sup>6</sup>

My tentative suggestion, then, based on some tentative intuitions about pairs of cases like these, is that some benefits of war, economic in one case and those of an unrelated peace in the other, can't help justify the war if they follow causally from a means to its ultimate just cause, but can if they follow from the just cause itself. That's a downstream causal distinction, like the one in Kamm's Principle and drawn on the same basis.

This isn't to say it's the same distinction. It concerns the morality of a whole war rather than of an individual act, though I imagine parallel cases can be constructed about individual acts. More importantly, it involves the production of goods rather than of evils. Kamm's principle concerns how the evil of the one death is caused: it's when that evil results from a means to the good of the five being saved that saving the five is forbidden, and when the evil results from the good itself that it's permitted. But my suggestion concerns the production of goods. The question is whether, for example, an economic benefit results from a means rather than from an end that determines whether that benefit can help justify a war. And in non-consequentialist moral views the causation of goods is often treated very differently from the causation of evils. Thus, many such views say it's morally more objectionable to actively cause an evil than merely allow the evil to come about, but they don't say it's morally more creditable to actively cause a good. If you can choose between actively saving one person and letting someone else save five, you should let the other save the five. So the fact that a certain distinction looks or may be important when we're producing goods doesn't imply that it's similarly important when we're producing evils; it may be relevant in one of those contexts but not the other.

Nonetheless, my tentative suggestion is in the same family as Kamm's, because it too draws a downstream distinction in how a morally significant result is caused, and draws it on the same basis, namely whether the result follows from something good or only from a means to it. The suggestion that this kind of distinction may matter in the moral assessment of actions was to my knowledge first made explicitly by Kamm and is among her many novel contributions to ethics. Even if it doesn't completely solve the Trolley Problem – and let's be clear that no one else has done that, in the sense of providing a satisfying justification for our initial intuitive judgements – it's a fruitful distinction to have in mind when exploring the intricacies of non-consequentialist ethics that have been the primary focus of Kamm's brilliant philosophical writings.

## Notes

1. F. M. Kamm, *Ethics for Enemies: Terror, Torture, and War* (Oxford: Oxford University Press, 2011), p. 142.
2. Judith Jarvis Thomson, "Killing, Letting Die, and the Trolley Problem," *The Monist* 59 (1976): 204-17.
3. Kamm, *Ethics for Enemies*, p. 142.
4. Kamm has highlighted the special moral status of neutrals in war, higher even than that of our noncombatants, in "Failures of Just War Theory: Terror, Harm, and Justice," *Ethics* 114 (2004): 650-92, pp. 672-73.
5. See my "Proportionality in the Morality of War," *Philosophy and Public Affairs* 33 (2005): 34-66, p. 40.
6. I discussed these two pairs of examples in "Proportionality and Necessity," in Larry May, ed., *War: Essays in Political Philosophy* (Cambridge: Cambridge University Press, 2008), 127-44, pp. 133-34.