Utilitas

Additional services for **Utilitas:**

On the Equivalence of Trolleys and Transplants: The Lack of Intrinsic Difference between 'Collateral Damage' and Intended Harm

HOWARD NYE

Link to this article: http://journals.cambridge.org/abstract_S0953820814000181

How to cite this article:
HOWARD NYE (2014). On the Equivalence of Trolleys and Transplants: The Lack of Intrinsic Difference between 'Collateral Damage' and Intended Harm. Utilitas, 26, pp 432-479 doi:10.1017/S0953820814000181

Request Permissions : Click here

On the Equivalence of Trolleys and Transplants: The Lack of Intrinsic Difference between 'Collateral Damage' and Intended Harm

HOWARD NYE

University of Alberta

In this article I attempt to show conclusively that the apparent intrinsic difference between causing collateral damage and directly attacking innocents is an illusion. I show how eleven morally irrelevant alterations can transform an apparently permissible case of harming as a side-effect into an apparently impermissible case of harming as a means. The alterations are as obviously irrelevant as the victims' skin colour, and consistently treating them as relevant would have unacceptable implications for choices between more and less harmful ways of securing greater goods. This shows not only how the principles philosophers have proposed for distinguishing between these cases cannot withstand scrutiny, but how we can be sure that there are no relevant differences yet to be discovered. I conclude by considering reasons to think that there are deontological constraints against harming, but that they apply just as forcefully against collateral harms as they do against intended harms.

I. INTRODUCTION

There is often an enormous moral difference between acts that inflict harm with the intention of inflicting it and acts that inflict harm without such an intention. Prompted by anger or a misguided desire for revenge, direct attacks on innocents may be intended to harm them as an end in itself, and lack justification entirely. Any benefits that might actually justify the harm of a direct attack, like those flowing from demoralizing an enemy, are guaranteed to be less likely to materialize than the harms one is certain to inflict if the attack succeeds.[1] By contrast, harms we inflict despite our best efforts to avoid inflicting them are often dramatically less likely to occur than the benefits that would justify the risk of inflicting them.[2] This is surely why drivers who scrupulously check their brakes seem to act permissibly (given

---

[1] For evidence that terrorism is generally ineffective see Max Abrahms, 'Why Terrorism Does Not Work', *International Security* 31 (2006), pp. 42–78. (But for evidence that it can under certain circumstances be effective see Eric Gould and Esteban Klor, 'Does Terrorism Work?', *The Quarterly Journal of Economics* 125 (2010), pp. 1459–510.)

[2] For instance, Gregory McNeal reports that the US military's 'collateral damage methodology' for pre-planned operations involving air-to-surface weapons and artillery seeks to ensure a probability of less than 1/10 of serious or lethal wounds to non-

their evidence), even on those extremely improbable occasions when, through no fault of their own, their brakes fail and they kill someone.

Is there, however, an *intrinsic* moral difference between directly attacking someone and harming her as a side-effect, distinct from the radically different probabilities of harm and benefit that such acts often have? The question has far-reaching implications for both ethical theory and practice, and has been dramatized by 'the trolley problem'. Consider the cases:

> *Transplant*. You are a surgeon, and the only way for you to save five children dying from organ failure is for you to remove the vital organs of one other child, Bugsy, and transplant them into the five. This will of course lead to Bugsy's death.

> *Trolley*. A trolley is headed towards five children tied to a track. You are standing by the track, and the only way for you to prevent the trolley from killing the five is to pull a switch that will divert it onto a side-track where it will kill Bugsy.

According to the intuitions of the vast majority of people, it is wrong to transplant Bugsy's organs into the five, but permissible to divert the trolley from hitting the five to hitting Bugsy.[3] Since in stylized cases of this kind we assume that you know the consequences of your acts with certainty, these intuitions – if they have probative force – suggest that there is an intrinsic moral difference between inflicting harm on someone as a means of benefiting others (in Transplant) and inflicting harm on someone as a side-effect of doing so (in Trolley).

'The trolley problem' is typically understood as that of giving a theoretical defence of the intuitive moral difference between cases like Trolley and cases like Transplant. There are, however, reasons to believe that this problem is unsolvable, because the intuitive difference between the cases cannot withstand scrutiny. If the salvation of five individuals can justify inflicting harm on one as a 'side-effect', we would

---

combatants, and in practice has assured that less than 1 per cent of such operations resulted in collateral damage. See McNeal, 'The U.S. Practice of Collateral Damage Estimation and Mitigation' (Pepperdine Working Paper, 2011), and 'Targeted Killing and Accountability', *Georgetown Law Journal* 102 (2014), pp. 681–794, at 730–54.

[3] John Mikhail reports that, of a diverse group of subjects, 92 per cent say it's wrong to perform the transplant and 90 per cent say it's permissible to divert the trolley. See Mikhail, 'Universal Moral Grammar: Theory, Evidence, and the Future', *Trends in Cognitive Science* 11 (2007), pp. 143–52. The trolley problem originates with Judith Thomson, building on a discussion of Philippa Foot. See Thomson, 'Killing, Letting Die, and the Trolley Problem', *The Monist* 59 (1976), pp. 204–17; Thomson, 'The Trolley Problem', *Yale Law Journal* 94 (1985), pp. 1395–415; and Foot, 'The Problem of Abortion and the Doctrine of Double Effect', *Oxford Review* 5 (1967), reprinted in her *Virtues and Vices* (Oxford, 2002), pp. 19–31, at 23 and 27–8.

seem justified in driving over Bugsy if this was the only way to get to the five in time to rescue them from drowning.[4] But this seems about as abhorrent as harvesting Bugsy's organs in Transplant. Attempts have been made to articulate principles that would distinguish between Trolley and Transplant without permitting us to run over Bugsy to save the five, but for reasons I will explain these principles are deeply implausible. We should take seriously the possibility that there is no morally relevant difference between the cases, and one of our intuitions about them is mistaken.

One possibility is that our intuition about Transplant is mistaken. We might initially think that harvesting Bugsy's organs is forbidden by deontological constraints against inflicting harm.[5] But if we cannot defend a distinction between harvesting Bugsy's organs and diverting a trolley into him, and we are unwilling to say that diverting is forbidden by deontological constraints, we seem forced to conclude that there are no deontological constraints against harvesting Bugsy's organs either. Perhaps we believe in such constraints only because the expected consequences of inflicting harm tend to be worse than those of failing to prevent harm, and the harms we inflict tend to be more salient than the harms we prevent by inflicting them. Make the harm to the five salient and the benefits to them certain, and you get a case like Trolley, where we intuit that it's permissible to inflict harm on Bugsy.

Another possibility is that our intuition about Trolley is mistaken. If we find deontological constraints against harvesting Bugsy's organs independently credible, and we cannot defend a distinction between harvesting his organs and diverting the trolley into him, we seem forced to conclude that diverting is forbidden by deontological constraints as well. Perhaps we fail to apply the constraint against inflicting harm to diverting the trolley only because, as Peter Unger suggested, we 'projectively group' Bugsy with the five as 'having a trolley problem', and it is insufficiently salient that, but for our interference, Bugsy would have been fine. Make that fact salient, and we might well intuit

---

[4] This is Foot's 'Rescue II' case. See Foot, 'Killing and Letting Die', *Abortion: Moral and Legal Perspectives*, ed. J. L. Garfield and P. Hennessey (Amherst, 1984), pp. 177–85.

[5] That is, we might think (roughly) that moral reasons against inflicting harm are intrinsically stronger than moral reasons to prevent harm, where reasons against inflicting harm are of sufficiently greater strength that our reasons not to kill one decisively outweigh our reasons to save five. There are ways of believing in something like this without believing in deontological constraints. One might be a consequentialist who believes that it's intrinsically worse for the world to contain inflictions of harm than failures to prevent harm, or a virtue ethicist who believes that any fully morally admirable person would be more intrinsically averse to inflicting harm than allowing harm. But the dialectic of the debate over the trolley problem would be identical for these views: it would simply be a debate about which outcomes to value intrinsically or which people to morally admire.

that it is wrong to divert the trolley – if, for instance, the only way to divert it was to ram another trolley into it, which would cause both trolleys to roll down a hill and crush someone napping in a hammock in his yard.[6]

Either way, it seems to make an important difference to ethical theory and practice if there is no moral difference between inflicting harm as a means in Transplant and inflicting harm as a 'side-effect' in Trolley. If Transplant must be assimilated to Trolley, it would appear that there are no stringent deontological constraints against inflicting harm. Furthermore, this would put pressure on us to reject moral permissions to favour our own interests and special obligations to the near and dear. For, as Shelly Kagan observes, if there is no intrinsic difference between allowing and inflicting harm, moral permissions and special obligations would seem to permit and require not only allowing but indeed *inflicting* greater harms on strangers in order to confer lesser benefits on ourselves and our loved ones.[7] But it seems more repugnant to accept this than simply to abandon the ideas of agent-centred permissions and special obligations.

If, on the other hand, Trolley must be assimilated to Transplant, it would appear that deontological constraints against inflicting harm are more far-reaching than we sometimes assume. Since civilian casualties seem inevitable in modern warfare, Jeff McMahan has suggested that acknowledging a constraint against inflicting harm as a 'side-effect' would commit us to *de facto* pacifism. While the extremity of the position depends upon exactly how stringent we think deontological constraints are, the point remains that we may need to rethink the permissibility of many acts that foreseeably put innocents at risk.

In this article I attempt to give a conclusive argument that there is no morally relevant difference between inflicting harm as a means in cases like Transplant and inflicting it as a side-effect in cases like Trolley. Several authors have argued against existing attempts to defend distinctions between such cases and pointed to other intuitions which suggest that there are no relevant distinctions.[8] While such

---

[6] See Unger, *Living High and Letting Die* (New York, 1996), ch. 4, esp. pp. 96–101 (for a similar suggestion see Thomson, 'Turning the Trolley', *Philosophy and Public Affairs* 36 (2008), pp. 359–74, at 374). Unger himself thinks that projective grouping helps us reach more accurate moral judgements, and that 'projective separating', or seeing some people as '*not* having the same problem', causes moral distortions. But Unger offers no argument for this; he never even considers questioning the intuition that it's permissible to divert the trolley.

[7] See Kagan, 'Does Consequentialism Demand too Much? Recent Work on the Limits of Obligation', *Philosophy and Public Affairs* 13 (1984), pp. 239–54.

[8] See Kagan, *The Limits of Morality* (Oxford, 1989); J. M. Fischer, 'Tooley and the Trolley', *Philosophical Studies* 62 (1991), pp. 93–100; Fischer, 'Thoughts on the Trolley

arguments lend credence to the conclusion that there is no moral difference between inflicting harm in Transplant and Trolley, they do not rule out the possibility that a successful defence of the distinction has yet to be discovered. More recently, Judith Thomson has given a positive argument for assimilating Trolley to Transplant. Although I am sympathetic to her conclusions, I explain in section II why I believe her argument is also inconclusive.

After making some methodological observations in section III, I show in section IV how we can, by making eleven morally irrelevant alterations, go from the Trolley case to the Transplant case. The moral irrelevance of the alterations is demonstrated by the direct implausibility of the idea that they make a moral difference once we fully understand what they are and the manifest ludicrousness of treating them as relevant in our moral reasoning. By demonstrating the irrelevance of these alterations, I show not only how the principles philosophers have proposed for distinguishing between Trolley and Transplant cannot withstand scrutiny, but how we can be sure that the cases really are morally equivalent.

Having argued in effect that we must choose between abandoning deontological constraints on harming and extending them to cover harm inflicted as a 'side-effect', I conclude in section V with some thoughts about how to make the choice. While I do not attempt to resolve the issue, I identify two considerations that I believe are centrally important. First, because deontological constraints are directly plausible, the initial burden of proof is against those who think we should do away with them entirely. But second, the extension of deontological constraints to the infliction of harm as a 'side-effect' threatens to make our moral reasoning depend implausibly on who will be harmed if we do nothing. Since my sympathies happen to be with the extension rather than the elimination of deontological constraints, I briefly examine how one might respond to the second consideration, and to what extent an extended deontology requires *de facto* pacifism.

II. THOMSON'S ARGUMENT

Thomson has offered the following argument in favour of the conclusion that diverting in Trolley, like harvesting in Transplant, is wrong. In rough outline, she first considers:

> *Bystander's Three Options*. A trolley is heading towards five children tied to a track. You are yourself trapped on a side-track to the left,

Problem', *Ethics: Problems and Principles*, ed. J. M. Fischer and M. Ravizza (Fort Worth, 1992), pp. 308–17; Alexander Friedman, 'Minimizing Harm: Three Problems in Moral Theory' (PhD Dissertation, MIT, 2002).

and you can: (1) pull a switch to the left, in which case the trolley will divert onto the left side-track and kill you, (2) pull the switch to the right, in which case the trolley will divert onto a side-track to the right and kill Bugsy, or (3) refrain from pulling the switch, in which case the trolley will kill the five.

Second, Thomson hopes we will agree that option 2 is morally wrong in this case. Third, she observes that it is directly implausible that the wrongness of option 2 depends on the presence of option 1. So if we judge it wrong to divert the trolley into Bugsy when we have the option of diverting it into ourselves, we must judge it wrong to divert it into Bugsy in the original Trolley case where we lack this option.[9]

At first sight it is unclear why Thomson's reader would agree with her that option 2 is impermissible in the three-option case. When I first read Thomson's paper I was convinced that diverting in Trolley was permissible, and her three-option case seemed to me equivalent to:

> *Bazooka Holder's Three Options*. You, Bugsy and the five are tied to three parallel trolley tracks, and empty trolleys are heading down each track. You have a bazooka with only two rounds, giving you three options: (1) blow up the trolleys heading towards the five and Bugsy, in which case only you will die; (2) blow up the trolleys heading towards the five and you, in which case only Bugsy will die; or (3) blow up the trolleys heading towards Bugsy and you, in which case only the five will die.

Here it seems clearly permissible to take option 2, even though this results in Bugsy's death. It would certainly be nice of you to take option 1, but, since there are no deontological constraints against simply saving one group rather than another, option 1 would be supererogatory. In the same way, if we think that there are no deontological constraints against diverting a trolley into one group rather than another, why should we think it wrong to divert the trolley into Bugsy rather than yourself?[10] Why not say that, while diverting the trolley into yourself would certainly be nicer than diverting it into Bugsy, this second option is still permissible?

---

[9] Thomson, 'Turning the Trolley', pp. 364–7. On p. 366 Thomson presents the third step by saying that unless we would be willing to divert the trolley into ourselves in the three-option case, we cannot decently regard ourselves as entitled to divert it into Bugsy in the two-option case. But on p. 367 she makes it clear that the argument is not supposed to depend on what we are willing to do. While Thomson never explicitly explains how the third step is supposed to go given this clarification, it seems that she is relying on what Unger calls a 'second-order intuition' that it's just obvious that it shouldn't matter to the permissibility of choosing option 2 whether option 1 is present.

[10] Cf. William Fitzpatrick, 'Thomson's Turnabout on the Trolley', *Analysis* 69 (2009), pp. 636–43, at 641.

Against this way of thinking about the case, Thomson suggests that saving the five would be a good deed, and appeals to the directly plausible idea that it is wrong to make others pay the costs of your good deeds instead of paying them yourself. By way of analogy she suggests that it would be wrong to donate to Oxfam by stealing money from others rather than sending one's own money.[11] In categorizing the diversion of the trolley away from the five as 'a good deed', Thomson is emphasizing the plausible idea that neither you nor Bugsy are morally required to divert the trolley into yourselves in order to save the five. It is indeed plausible that it is wrong to force a cost on someone if she isn't morally obligated to assume it. There are, however, cases in which we *do* seem permitted to do things that result in someone's bearing a cost that she would not be obligated to assume herself. We seem permitted, in cases like Bazooka Holder's Three Options, to save ourselves rather than Bugsy even though Bugsy would not be required to save us rather than himself.

What really lends force to Thomson's claim is its invocation of deontological constraints against inflicting harm. The idea isn't that it's wrong for us to act in any way that *results* in someone bearing a cost she isn't obligated to assume voluntarily. The idea is that it's wrong for us to *inflict* a cost on someone that she isn't obligated to assume. This is a plausible idea, and it supports Thomson's conclusion. But the problem is that proponents of the idea that it is permissible to divert the trolley in the standard two-option case are committed to the view that, if there are any deontological constraints against inflicting harm, they do not apply to cases of diverting it or inflicting it as a 'mere side-effect'. These proponents will not be impressed by the analogy to stealing to give to Oxfam (or more plausibly: harvesting someone else's organs rather than giving up your own) because, if they think this wrong, they will see it as an instance of the 'problematic' kind of infliction of harm that goes on in Transplant, rather than the 'unproblematic' kind of infliction that goes on in Trolley.[12]

Thus, while Thomson's argument brings out some intuitive ideas that conflict with the intuition that it's permissible to divert in Trolley,

[11] Thomson, 'Turning the Trolley', pp. 364–5. Cf. Unger's 'Principle of Ethical Integrity' in his *Living High*, pp. 139–43. Unger, however, seeks to argue from the intuitive permissibility of imposing losses on others by diverting trolleys to the claim that giving to Oxfam is no mere good deed, in that failing to do so is seriously morally wrong. Since earlier in his book Unger (following Peter Singer, 'Famine, Affluence, and Morality', *Philosophy and Public Affairs* 1 (1972), pp. 229–43) makes an extremely strong independent case that we are morally required to donate most of our resources to organizations like Oxfam, Thomson's analogy is less than ideal.

[12] See Fitzpatrick, 'Thomson's Turnabout', pp. 638–9 and Thomson, 'Killing, Letting Die, and the Trolley Problem'.

philosophical defenders of a moral difference between Trolley and Transplant seem to have an escape-route. They can deny her intuitions about the three-option case and insist, as Thomson herself once did, that the plausible idea that we can't inflict costs on people who aren't obligated to assume them doesn't really apply to cases like Trolley. I believe that the only way to make the argument conclusive is to show that there is no place between Trolley and Transplant for these friends of a moral difference to hide.

III. METHODOLOGICAL CONSIDERATIONS

The mere fact that a moral theory fits our initial intuitions about what's wrong or permissible to do in particular cases cannot be a sufficient reason to accept it. For any pattern of intuitions whatever, we can find some gerrymandered moral theory to fit it. If we were justified in accepting any theory that fits our case-intuitions, a racist could consult his intuitions and justifiably conclude that, initial appearances to the contrary, someone's facial features, skin colour and continent of ancestry must make an intrinsic difference to how much one should care about her interests. While racist intuitions are parochial, they reflect a universal tendency to care more about those who share with us any traits that context makes salient.[13] The principle 'do more to help those who share with you any trait that has been raised to salience' would probably fit unreflective intuitions nicely, but is surely unacceptable.

Of course, as its name attests, we view this tendency to favour members of a salient in-group as a bias, so we view the intuitions it generates as tainted. But how do we know that the tendency is a bias, as opposed to a virtuous propensity to 'take care of our own'? Surely because it just *seems obvious* that sheer salience, like whether it is Tuesday or what colour someone's eyes happen to be, can be of no intrinsic moral importance. By contrast, it seems obvious in the abstract that the fact that an act will promote someone else's well-being *is* a moral reason to perform it, so we should not treat the sensitivity of intuitions about an act's wrongness to its effects on the well-being of others as a bias. In sorting out which case-intuitions to trust, we must determine which can be supported by general moral ideas that are *directly plausible*, or seem *a priori* to be true independent of what else entails them and what else they entail.[14]

---

[13] For a review see Marilynn Brewer, 'In-Group Bias in the Minimal Intergroup Situation', *Psychological Bulletin* 86 (1979), pp. 307–24. Psychologists have documented in-group bias in response to such trivial similarities as a shared tendency to over- or under-estimate the number of dots flashed on a screen.

[14] See Singer, 'Sidgwick and Reflective Equilibrium', *Monist* 58 (1974), pp. 490–517, at 515–17.

We cannot, however, rest content with initial appearances of plausibility, since the idea that a factor makes a moral difference might prove implausible on closer inspection. The racist injunction to 'take care of one's own kind' might sound plausible at first, but when we clarify that to be of 'one's own kind' is simply to share features like skin colour and area of ancestry, the plausibility of the idea vanishes. To determine whether a factor is morally relevant in its own right, we must clarify what it really amounts to and what it would be like consistently to treat it as relevant in our moral reasoning.

In what follows I argue that there are no morally relevant factors that separate inflicting harm in cases like Trolley from inflicting harm in cases like Transplant. I show that we can go from Trolley to Transplant by making eleven alterations in the underlying factors, none of which are morally relevant. I submit that each of the eleven alterations is, like salience and skin colour, irrelevant on its face. While philosophers have proposed several moral distinctions among the cases I consider, I argue that, on reflection, it seems obvious that these distinctions make no moral difference.

Of course, even if when carefully clarified the principles to which philosophers have appealed to justify the relevance of the factors separating Trolley from Transplant are deeply implausible, I cannot anticipate and respond to all possible attempts to derive the non-instrumental significance of the factors from further principles. No ethical argument can be 'conclusive' in the sense of anticipating and responding to all possible arguments against its conclusion. The sense in which I take my argument to be conclusive is that it shows the only factors separating Trolley from Transplant to be as obviously devoid of *basic* significance, and as unlikely to be of derivative non-instrumental significance, as factors like skin colour and salience.[15]

I argue in particular that drawing a moral boundary at any of the eleven steps I consider would have unacceptable implications for choices between more and less costly ways of preventing five deaths. When there is no deontological constraint against saving five by doing

[15] I believe that once we fully appreciate what the differences between the cases amount to and how irrelevant they look, it is clear not only that they cannot be of underivative moral significance, but that there is just as little hope of deriving their non-instrumental significance from distinct factors as there is of doing so for skin color and salience. (There are certainly differences which are of derivative *instrumental* significance – such as differences which make the benefits of harvesting in Transplant less certain in the actual world than those of diverting in Trolley – but these are not at issue.) This is what I mean when I claim that these factors are obviously devoid of intrinsic moral significance. I am grateful to a referee from *Utilitas* for raising the issue of derivative non-instrumental significance, and for encouraging me to clarify the sense in which I take my argument to be 'conclusive' (or at least more conclusive than its predecessors).

something that results in one dying, it is permissible to engage in that kind of conduct even if it results in two (or indeed four) dying, if that is the only way to save five. Thus, it is permissible not only to save five rather than one but to save five rather than two, and if there is no constraint against diverting the trolley from five into one, it is surely also permissible to divert it from five into two. Consider, then, a three-option choice among: (1) saving five by doing something that *isn't* prohibited by a deontological constraint, which results in *two* individuals dying, (2) saving five by doing something that *is* prohibited by a deontological constraint, which results in only *one* (different) individual dying, and (3) not saving five. Since simply adding option 1 to the choice between options 2 and 3 obviously cannot remove the prohibition against option 2 or introduce a prohibition against option 1, option 2 must remain wrong, while option 1 must remain permissible and favoured over option 3 by the fact that it will save more lives.[16] Consider, for instance:

> *Organ Failures*. There are seven children about to suffer organ failure: five to your left and two to your right. You have drugs that can prevent organ failure, but only if administered immediately, and you cannot reach both the five and the two in time. You are, however, also a transplant surgeon, and Bugsy, whose organs are a match for the five (but not the two) is waiting for a routine check-up back at the hospital, giving you three options: (1) rush to the five, in which case only the two will die, (2) rush to the two, then harvest Bugsy's organs and transplant them into the five, in which case only Bugsy will die, or (3) rush to the two, then leave Bugsy alone, in which case only the five will die.

If, as seems plausible, there is a deontological constraint against inflicting harm but not against simply failing to prevent harm, you should choose option 1 over option 2 because option 2 is prohibited by that deontological constraint while option 1 is not. You can on this view justify taking option 1 over option 2 by pointing out that option 2 would have killed Bugsy while option 1 simply saved the five rather than the two. The fact that this *does* seem to be a perfectly good justification for taking option 1 supports the plausibility of a deontological constraint against inflicting harm.

---

[16] Cf. Unger, *Living High*, p. 93. It is important that the individual who dies if you take option 2 is distinct from those who die if you take option 1, since otherwise we might plausibly reason that 'we are already allowed to leave him badly off, so it can't be wrong to do something to him that leaves him no worse off' (see Frances Kamm, *Intricate Ethics* (New York, 2007), pp. 169–73, which makes trouble for Friedman's arguments from multiple option cases in his 'Minimizing Harm').

If a moral boundary could be drawn at any of the eleven alterations between Trolley and Transplant that I consider, then killing one to save five in cases on one side of the boundary would have a feature in virtue of which it is prohibited by a deontological constraint, while killing one to save five in cases on the other side of the boundary would lack this feature. Thus, if we had a choice between (1) saving five by killing two in a way that lacks the problematic feature, (2) saving five by killing one in a way that has the problematic feature, or (3) not saving five, we should choose option 1 over option 2. For each candidate feature I show how this conclusion would be indefensible.

One might worry that in such three-option cases the salience of killing in both options 1 and 2 obscures a factor in virtue of which only option 2 is prohibited by a constraint, leading us to conclude erroneously that we should take option 2 over option 1 to minimize harm. But the lack of similar obscuring in Organ Failures makes it dubious that plausible grounds of constraints can be so easily obscured in three-option cases. In Organ Failures both options 1 and 2 have the salient property of saving more lives, but this neither obscures the fact that option 2 involves inflicting harm, in virtue of which (plausibly) only it is prohibited by a constraint, nor inclines those who find constraints plausible to conclude that we should take option 2 over option 1 to minimize harm. Moreover and much more importantly, I do *not* rely in my argument on unsupported intuitions about what it is permissible to do in three-option cases. I rely instead on whether it is directly plausible or implausible to *cite* specific differences between options 1 and 2 as *justifications* for taking option 1 once we fully appreciate what the differences amount to. This prevents the differences from being obscured and allows us to assess whether they really are important enough to justify acting in ways that result in fewer individuals surviving.[17]

It is, however, just possible that there is no single relevant difference between Trolley and Transplant. It is conceivable, if quite implausible, that there are several differences between the cases, none of which makes all the difference on its own, but which cumulatively make it permissible to kill one to save five in Trolley but not Transplant. Since I am considering eleven differences between the cases, for such accumulation to work at least 1/11 of the moral difference must

---

[17] I am indebted to a referee from *Utilitas* for raising the objection that multiple-option cases might obscure the relevance of certain factors, which I believe is a real problem with Unger's reliance on unsupported intuitions about permissibility in such cases. My focus on the plausibility or implausibility of citing factors as justifications (once we understand what they amount to) is a respect in which I take my argument to be more conclusive than Unger's.

manifest itself at some step I am considering. I employ two conservative ways of thinking about whether a factor makes at least 1/11 of this moral difference: (i) asking whether it justifies taking an additional 1/11 risk of killing someone, and (ii) asking whether it justifies deducting an additional seven years from the seventy-seven further years that each of the children in my cases will live if they survive unscathed.[18] In what follows I argue that there is *no* relevant difference between *any* of the cases ranging from Trolley to Transplant, and I conclude that the cases are morally equivalent.[19]

IV. AN ELEVEN-STEP PROGRAMME FOR TROLLEY INTUITIONS

Recall the original Trolley case, in which the only way to save five children tied to a track is to divert the trolley about to hit them into

---

[18] This strategy is doubly conservative. First, since the whole moral difference would have to be distributed among the eleven steps, if less than 1/11 of the difference were present at *any* step, *more* would have to be present at others. Second, if we subscribe to the Prioritarian idea that we should be more concerned with decreases in well-being the worse off the individual suffering them will be (see Derek Parfit, 'Equality and Priority', *Ratio* 10 (1997), pp. 202–21), then decreasing someone's future years from seventy-seven to seventy may have *less* than 1/11 of the moral importance of decreasing someone's future years from seventy-seven to zero.

[19] One might worry that my argument has the form of a sorites (a bazillion grains of sand make a heap, the subtraction of a single grain from a heap can't make a non-heap, so 1 grain of sand is a heap), and is unsound for whatever puzzling reasons sorites arguments are unsound. But my argument involves no continuous dimension of variation, and eleven steps seem far too few to generate a sorites. To the extent that there is a sorites-esque worry, it seems to be that several small moral differences between Trolley and Transplant are diffusely distributed among the steps I consider, which I address by examining whether 1/11 of the difference is present at any step.

While my broader aim is to establish the stronger conclusion that there are no intrinsic moral differences *of any kind* between diverting in Trolley and harvesting in Transplant, my tests for 1/11 of the difference between permissibility and wrongness most directly support the weaker conclusion that there are no intrinsic moral differences *which make diverting permissible but harvesting wrong*. One might be inclined to accept only the weaker conclusion if one thought some differences between the acts were intrinsically important, but insufficient to make the difference between permissibility and impermissibility. To test for this, I invite the reader to consider whether any of the differences I discuss could justify any additional risk or harm *less than* a 1/11 chance of death or deprivation of seven years. One might also be inclined to accept only the weaker conclusion if one thought that the salvation of five lives was insufficient to justify either diverting in Trolley or harvesting in Transplant, but that more good would be needed to justify harvesting Bugsy's organs than diverting a trolley into him. To test for this, I invite the reader to increase the number of children who could be saved by inflicting harm in the cases under consideration from five to ten, one hundred, or whatever large number one wishes (see for instance my discussion of using this test to argue against certain of Warren Quinn's views in n. 51). I submit that my remarks below, reinterpreted to apply to these tests, support the stronger conclusion. I am grateful to a referee for *Utilitas* for drawing my attention to the importance of inviting readers sceptical of the stronger conclusion to consider these alternative tests.

Bugsy. Our first step is to ask whether it would matter if diverting the trolley caused a *different* trolley to hit Bugsy, as in:

> *Shifted Track*. The only way to prevent a trolley from killing five children is to pull a switch that will remove a section of track from its path, which will cause it to swerve away harmlessly. Unfortunately, pulling the switch will also cause the section of track to connect the track of a second oncoming trolley to a track to which Bugsy is tied, and rather than swerving away harmlessly this trolley will kill Bugsy.

It seems obvious that there is no moral difference between diverting the trolley in Trolley and Shifted Track. Surely it cannot matter whether diverting it kills Bugsy with the same trolley or a different one.

Some authors have, however, suggested that it is easier to justify inflicting harm by 'diverting an existing threat' than by 'creating a new threat', and it might seem that pulling the switch in Trolley merely 'diverts an existing threat' while doing so in Shifted Track 'creates a new one'. But if a 'threat' is anything that will participate in an event or interaction that will cause harm, the threats to the five include not only the trolley, but the section of track that will guide the trolley into them. Pulling the switch in Shifted Track transfers the track-section from threatening the five with a trolley to threatening Bugsy with a trolley, and in this sense 'diverts an existing threat'.

The only way to restrict the threats to the trolleys would be to construe 'threats' as something like things that will 'directly' cause harm, say by harmfully interacting with someone's body. But on this reading, you wouldn't count as diverting an existing threat in either Shifted Track or Trolley if there were stationary trolleys parked in front of Bugsy and the five, and the oncoming trolleys will kill by hitting these stationary trolleys and causing them to crush whoever is on the other side. Since it obviously cannot matter morally whether the trolleys themselves run over victims or smash into other trolleys that do this, it cannot matter morally whether you kill someone by 'diverting a "direct" threat' or 'creating a new "direct" threat'.[20] Proponents of a moral distinction between diverting and creating threats find it plausible that it is relatively easy to justify 'arranging that something that will

---

[20] If you think it matters that the parked trolleys be not merely overdetermining causes of the deaths of the victims (cf. Kamm, *Intricate Ethics*, pp. 143–4, 159–60), suppose that the oncoming trolleys lack sufficient momentum to crush the victims, but kill by creeping onto and depressing platforms located directly in front of the parked trolleys, which send the parked trolleys forward over the victims. Obviously this too makes no moral difference.

do harm anyway . . . shall do harm to fewer rather than more'.[21] But whether the thing in question does harm by interacting with victims' bodies or causing other things to do so seems as obviously devoid of intrinsic moral significance as skin colour.

To confirm the lack of moral difference between diverting the trolley in Trolley and Shifted Track, suppose that you had three options: (1) divert the trolley headed towards the five onto a track with both Freddy and Suzy, (2) pull a switch that moves a track-section from connecting the trolley to the five to connecting a second trolley to Bugsy, or (3) do neither 1 nor 2. It seems indefensible to choose option 1 over option 2. How could you explain to Freddy and Suzy's parents why you killed both of them rather than only Bugsy, or no one at all? Or suppose that Suzy wasn't squarely centred on the track, so that in addition to killing Freddy, option 1 merely ran a 1/11 risk of killing Suzy, or injured her so badly that she dies seven years earlier than she otherwise would. Some fumbling suggestion about diversions of direct versus indirect threats is clearly an insufficient justification for killing an additional child, exposing her to a 1/11 risk of death, or depriving her of seven years of life.

Our second step asks whether it would matter if you connected and disconnected the tracks by means of different intermediaries, as in:

> *Different Sections*. The only way to prevent a trolley from killing five children is to push a button that will simultaneously throw two relay switches. The first switch will remove a section of the trolley's track, which will cause it to swerve away harmlessly. Unfortunately the second switch will cause a different section of track to connect a second oncoming trolley to a track to which Bugsy is tied, and rather than swerving away harmlessly it will kill Bugsy.

It seems obvious that there is no moral difference between disconnecting the first track and connecting the second in Shifted Track and Different Sections. Surely it cannot matter whether you do this with a single track-section or with relays that move different track-sections.

Pushing the button in Different Sections, unlike shifting the single track-section, does not seem to 'divert' anything from causing harm to the five to causing harm to Bugsy. But why should it matter whether, in benefiting the five and harming Bugsy, you shift some of the very same things from causing the deaths of the five to causing Bugsy's death instead? Some authors have characterized things that can be

---

[21] Thomson, 'The Trolley Problem', p. 1408.

shifted from participating in interactions that will harm the five to participating in interactions that will harm Bugsy as 'generalized perils' or 'public threats'.[22] But this falsely suggests that absent your intervention these things 'threaten' or will cause harm to Bugsy as well as the five. Thomson once suggested that we should view transferable causes of death as 'bad [things], up for distribution'.[23] But if we are permitted to 'redistribute' the harm of death from the five to Bugsy by making a track-section connecting a trolley to the five instead connect a trolley to Bugsy, why should we not be equally permitted to redistribute this harm by pressing a button that removes this track-section and connects a trolley to Bugsy with a different track-section? If, alternatively, we are not permitted to 'redistribute' death by pressing a button that connects a trolley to Bugsy and disconnects a trolley from the five by moving different track-sections, why should we view a single track-section as 'up for distribution' in a sense that permits us to shift it in order to achieve this very same result?[24]

On reflection, it seems directly implausible that transference of the selfsame physical object should make an intrinsic moral difference. Suppose we initially thought that flipping a switch would shift a single track-section from connecting a trolley to the five to connecting a trolley to Bugsy. Suppose we were about to flip the switch, when a proponent of the diverting–creating distinction burst onto the scene with allegedly urgent news: 'Wait! Stop! If you flip the switch it won't transfer the *same* track-section from connecting a trolley to the five to connecting a trolley to Bugsy! Flipping the switch will cause a *totally different* track-section to connect a trolley to Bugsy!' Surely it would be absurd for us to change our minds about whether to flip the switch in response to this information.

Frances Kamm has suggested a second moral difference between Shifted Track and Different Sections. According to Kamm, when you divert in Trolley, the events by which you harm Bugsy (like the trolley

[22] Eric Mack, 'Three Ways to Kill Innocent Bystanders', *Social Philosophy and Policy* 3 (1985), pp. 1–26, at 17; and Fitzpatrick, 'Thomson's Turnabout'. See also James Montmarquet, 'On Doing Good: The Right and the Wrong Way', *The Journal of Philosophy* 79 (1982), pp. 439–55, at 446–55; and Robert Hanna, 'Morality De Re', *Ethics: Problems and Principles*, ed. J. M. Fischer and M. Ravizza (Fort Worth, 1992), pp. 318–36, at 328–32.

[23] Thomson, 'Killing, Letting Die, and the Trolley Problem', p. 215.

[24] Michael Costa claims that 'redirecting an existing threat . . . is not . . . something that is evil in and of itself' but 'bring[ing] some new threat or source of foreseeable harm into the world . . . is itself a prima facie evil' (Costa, 'Another Trip on the Trolley', *Southern Journal of Philosophy* 25 (1987), pp. 461–6, at 464). But surely neither pulling a switch nor pushing a button is 'evil in and of itself' in the standard sense of 'independent of its consequences', and it is entirely unclear in what sense or why one of these acts but not the other would be 'prima facie evil' when they both have the same consequences of disconnecting the first track and connecting the second.

changing direction) don't actually cause the greater good of the five being saved; they instead have this greater good as their 'non-causal flip side'. Of course, there are plenty of events that harm Bugsy and cause the five's salvation, like your arm moving, the switch moving, and electric current flowing. But these explain the salvation of the five only by explaining why the trolley moves. Kamm's idea seems to be that in Trolley (and Shifted Track), the event that *most proximately explains both* the good and the harm – the movement of the trolley (or track-section) – itself constitutes the five being free from threats. But in Different Sections the most-proximate-joint-explainer of harm and benefit (the depression of the button) does not constitute the freedom of the five from threats; it merely causes such freedom by causing the distinct event of the first track-section moving.[25]

On its face, the fact that an act inflicts harm on Bugsy and saves five by means of most-proximate-joint-explainers that constitute rather than cause the freedom of the five from threats looks as obviously devoid of intrinsic moral significance as skin colour. Kamm suggests, however, that the movement of the trolley in Trolley constitutes not only the freedom of the five from threats, but the greater good of the five being saved, and that this is morally important. It is, however, simply false that the movement of the trolley itself constitutes the benefits that count in favour of diverting it. The trolley's turning away from the five at time t benefits them by making it the case that rather than dying at $t + 1$, they live from $t + 1$ to $t + n$. But the benefit of their living from $t + 1$ to $t + n$ is by no means identical to or constituted by the trolley turning away from them at t.[26]

More importantly, the fact that a most-proximate-joint-explainer of a harm and a benefit constitutes rather than causes the benefit itself *also* seems about as morally relevant as skin colour. Kamm claims that, when you harm B and benefit F by causing a most-proximate-joint-explainer that *constitutes F*'s benefit, you merely *substitute B for F* as the party who is harmed, while if you harm B by causing a most-proximate-joint-explainer that *causes F*'s benefit, you *subordinate B to F* (164–6). 'Subordination' sounds morally worse than 'substitution', but Kamm appears to be affixing relevant-sounding labels to manifestly irrelevant metaphysical distinctions. In the same way that the initial

---

[25] Kamm, *Intricate Ethics*, pp. 140–4. On Kamm's view this difference might not forbid diverting in Different Sections, because the harm this causes might count as 'indirect' (see discussion of Dual-Action Bomber below).

[26] The events are totally separated from each other in time: one begins and ends at t, while the other begins later at $t + 1$ and lasts all the way to $t + n$. The events are also completely separated in space: one takes place on a bit of trolley track some way away from the five, while the second takes place in and around the five and everywhere they go (home, work, the park, etc.).

plausibility of the racist maxim 'take care of your own kind!' vanishes when we see that 'own kind' amounts to nothing more than shared skin colour, the plausibility of a Kammian distinction between 'substitution and subordination' seems to vanish when we see that it amounts to nothing more than a distinction between most proximate-joint-explainers that constitute rather than cause someone's being free from threats.

Kamm's attempt to connect her talk of substitution and subordination to directly plausible, non-stipulative ideas seems half-hearted and confused. Some of her remarks seem to conflate her distinction with others, like that between diverting and creating threats (discussed above) and that between harmful means that do and do not involve victims (discussed below).[27] The rest hinges on her claim that we must not divert a trolley from hitting a tool that can be used to save many people to hitting one person, or use such a tool when this will harm someone, because this would imply:

> That the many people are transmitting their importance to the tool; these people would have the power to change the value of the things in the world we all share . . . This elevation of the tool over a person, in virtue of its usefulness for others, results in the subordination of the one person to the other people. (165–6)

The idea seems to be that if your most proximate harmful means of saving some people cause rather constitute their freedom from threats, the means act as a tool to which those harmed are subordinated, and given this they are subordinated to those saved by the tool. Why, however, should it matter whether your *most proximate* harmful means cause harm? Why not say that in moving your arm or pulling the switch in Trolley you subordinate Bugsy to your arm or the switch, and in turn subordinate him to the five you save with these things?

---

[27] Kamm claims that 'when one person's not being threatened is the non-causal flip side of sending the threat elsewhere' anyone threatened 'occupies the very same position that another person would have occupied relative to the threat', so 'this is substitution' (165). But this applies only to a distinction between diverting threats and creating new ones. Kamm's distinction is supposed to apply more broadly; it is, for instance, supposed to entail that it's permissible to move five people away from a trolley, even though this causes them to tumble down a hill and crush one (Kamm's 'Tumble Case', 139). Kamm also claims that when we harm someone 'as a causal means to save someone else from a threat . . . the position he occupies . . . makes essential reference to his usefulness to achieving a good for that other person' (165). But it is simply untrue in many cases of inflicting harm that Kamm's distinction is supposed to condemn that the victim is herself a useful means of achieving goods. Take Foot's Rescue II, where the only way to save five drowning swimmers is to drive over one individual trapped in the road. Here the effect you have on the one does nothing to help you save the five – she is in no way useful to achieving goods for them.

   While talk of tools may lend superficial plausibility to the idea that worrisome subordination is present, suppose our choice was to *save* either Bugsy or a tool that we *knew* would save five who would otherwise die (say a needle with enough life-saving serum). Surely it would be permissible to save the tool, and preposterous that this involves some objectionable elevation of the tool or the five over Bugsy.[28] In choosing to save the tool we are not fetishizing it as *intrinsically* valuable; the only sense in which the five 'transmit their importance' to the tool is that we should care *instrumentally* for its salvation out of our intrinsic concern for them. To say that an instrumental preference for pulling the tool rather than Bugsy's body out of the water 'subordinates' Bugsy to the tool and thence the five seems equivalent to the manifestly absurd claim that choosing to row right towards five drowning swimmers instead of left towards one subordinates the one to rowing right, which subordinates him to the five. Exactly similar remarks go for diverting the trolley from the tool into Bugsy.[29]

   To confirm the irrelevance of these diverting-creating and Kammian distinctions, suppose that you actually had three options: (1) pull a switch that transfers a single track-section from connecting the first trolley to the five to connecting a second trolley to Freddy and Suzy, (2) push a button that activates two relay switches, one of which removes the track-section connecting the first trolley to the five, but the other of which connects a third trolley to Bugsy with a different track-section, or (3) do neither 1 nor 2. Surely it would be indefensible to choose option 1 over option 2. Could you really with a straight face try to justify your act to Freddy and Suzy's parents on the grounds that 'had I killed Bugsy instead of both Freddy and Suzy, it wouldn't have been with any of the things that would have killed the five', or 'the most proximate means by which I would have done both good and harm would have caused rather than constituted the five's being free

[28] John Taurek might complain that it is unfair to Bugsy that he be abandoned simply because he is in the less numerous group (Taurek, 'Should the Numbers Count?', *Philosophy and Public Affairs* 6 (1977), pp. 293–316). If you are moved by this thought then any time I speak of the permissibility of saving the five assume that I mean the permissibility of saving them after having flipped a fair coin to determine whom to save. There is still a debate to be had over the trolley problem, because intuitively it is not permissible to flip a coin to determine whether to harvest Bugsy's organs in Transplant (cf. Friedman, 'Minimizing Harm', pp. 162–3). For good reasons not to be moved by Taurek's thought, see e.g. Parfit, 'Innumerate Ethics', *Philosophy and Public Affairs* 7 (1978), pp. 285–301.

[29] Of course, if it is *for independent reasons* permissible to save the tool rather than Bugsy but wrong to divert the trolley from the tool into Bugsy (say because it's harder to justify inflicting harm than failing to prevent harm), we might truly say that in the first case but not the second you save the five *by wronging Bugsy*, and *in this sense* 'subordinate' him to the five. But whether you 'subordinate' someone in this sense is explained by, and cannot itself explain, whether you wrong her.

from threats'? These would be preposterous justifications for killing an additional child, or, if Suzy wasn't centred on the track, for exposing an additional child to a 1/11 risk of death or depriving her of seven years of life. Making baseless noises like 'Bugsy would have been subordinated while Freddy and Suzy were merely substituted' would not detract from the preposterousness.

Our third step asks whether it would matter if, instead of causing a moving trolley to hit Bugsy, the button caused an initially stationary trolley to hit him, as in:

> *Start-Up*. The only way to prevent a trolley from killing five children is to push a button that will throw two relay switches. The first switch will remove a section of the trolley's track, which will cause it to swerve away harmlessly. Unfortunately the second switch will cause an initially stationary second trolley to move along a track to which Bugsy is tied, which will kill him.

It seems obvious that there is no moral difference between pressing the button in Different Sections and Start-Up. Surely it cannot matter whether the button causes an initially moving or an initially stationary trolley to hit Bugsy.

Michael Moore has, however, suggested that it is permissible to save five by doing something that diverts a moving trolley into Bugsy because 'there is a big natural cause already in play' and 'your act makes a small [causal] contribution [to Bugsy's death] relative to [it]', while it would be wrong to save the five by doing something that sets a stationary trolley moving into Bugsy, because your act would make 'a larger causal contribution' to Bugsy's death. This is surprising, because as Unger pointed out, this sort of 'protophysical' distinction between altering a moving object's direction and setting an object at rest in motion seems 'not just ethically, but even physically . . . absolutely empty'.[30]

While relativization to the actor's frame of reference gives physical content to the distinction between moving and stationary trolleys, nothing will, I believe, make sense of Moore's attempt to distinguish morally between such cases in terms of whether your causal contribution to Bugsy's death is 'small'. First, suppose that out of sheer indifference to Bugsy's interests you diverted a trolley from harmlessly moving down an unoccupied main track onto the side-track on which Bugsy was standing. It would be absurd to claim that your causal contribution to Bugsy's death was 'small' and that 'you didn't do much'.

---

[30] See Moore, *Causation and Responsibility* (Oxford, 2009), pp. 75–6; Unger, *Living High*, p. 101.

But whatever the size of your causal contribution to Bugsy's death is supposed to be, it cannot be altered by the presence of the five on the main track. Second, suppose that to harvest Bugsy's organs and transplant them into the five, all you had to do was press a button that would divert an organ-harvesting robot from harmlessly moving down the hallway to entering Bugsy's room, where it will automatically remove his organs. It would be ridiculous to say that this – unlike wielding the scalpel yourself – is permissible because it 'wouldn't do much' and 'would make a small contribution' to Bugsy's death. Third, 'the size of your causal contribution to Bugsy's death' would presumably have to be something like the amount of conserved-quantity transfer or number of events you must cause to make something kill him. But suppose that in the original Trolley case there are massive forces holding the trolley on its course, and you must create enormous forces or Rube-Goldberg-like series of events to divert it. Suppose, however, that in Start-Up the second trolley is on frictionless, downward-sloping tracks, and the button starts its motion by giving it a tiny nudge. Now your causal contribution to Bugsy's death is greater if you divert the moving trolley than if you move the stationary trolley, but I doubt Moore wants to say this makes it wrong to divert the moving trolley but permissible to move the stationary one.

Whatever Moore wants, it is obvious that the amount of conserved-quantity transfer or number of events involved in killing someone is as devoid of intrinsic moral significance as skin colour. His baffling suggestion to the contrary seems to be a confused extension of the already dubious idea that it is easier to justify doing what you know will cause other agents to harm victims than it is to justify harming victims by means that bypass the decisions of others. When you cause others to harm victims, it can be tempting to 'put the blame on' the others, pleading that, after you acted, they could have chosen not to do harm. But it is *just crazy* to 'put the blame on' the trolley for killing Bugsy if you divert it into him; after you pulled the switch, the trolley most certainly couldn't have chosen not to run him over! Blaming the trolley instead of yourself for Bugsy's death clearly requires you to ignore the fact that absent your interference it would not have harmed him.[31]

To confirm the irrelevance of Moore's distinction, suppose you had three options: (1) push one button that disconnects the first trolley

---

[31] Perhaps you can 'blame the trolley' for its being true that *someone* will die in Trolley, but you can just as easily 'blame the trolleys and the effects of the button' for the fact that someone will die in Different Sections, or even 'blame the organ failures and the inability of Bugsy and the five to live without organs' for the fact that someone will die in Transplant.

from the five and connects a second moving trolley to Freddy and Suzy, (2) push another button that disconnects the first trolley from the five and sends an initially stationary third trolley into Bugsy, or (3) do neither 1 nor 2. It would be indefensible to choose option 1 over option 2. You could not seriously try to justify this to Freddy and Suzy's parents by saying 'my causal contribution to Bugsy's death would have been big, but my causal contribution to Freddy and Suzy's deaths was small'. This sounds false, and would in any event be an absurd justification for killing an additional child or (if Suzy wasn't centred on the track) for exposing an additional child to a 1/11 risk of death or depriving her of seven years of life. Adding 'I would have been to blame for Bugsy's death, but the second trolley is to blame for Freddy and Suzy's death' would be *outrageously* absurd.

Our fourth step asks whether it would matter if the button that saves the five caused bombs to fall on Bugsy, as in:

> *Dual-Action Bomber*. The only way to prevent a missile from killing five children is to press a button that will drop the bombs on the front of your plane onto the missile. Unfortunately this button will also drop the bombs on the rear of your plane, which you know will land on Bugsy's house. Bugsy is in his basement under reinforced concrete that would ordinarily protect him, but vibrations from the impact of the bombs will ignite chemicals in his basement that will explode and kill him.

It seems obvious that there is no moral difference between pressing the button in Start-Up and Dual-Action Bomber. It cannot matter that the button saves five and kills Bugsy through explosions instead of trolley movements.

One *might* suggest that because the bomber and the bombs, but not the trolleys, were constructed in order to destroy things, they have this as 'their function', so it is somehow 'less of an accident' if the rear bombs kill Bugsy in this case than if the second trolley kills him in Start-Up.[32] But surely it makes no intrinsic moral difference what anyone's intentions were in making the trolleys and bombs – it wouldn't change anything if we discovered that the trolleys were constructed for the purpose of running people over. One might also suggest that saving the five in Dual-Action Bomber kills Bugsy by 'introducing a threat (your bombs) into his environment', while doing so in Start-Up kills him with

---

[32] This might be suggested by Matthew Hanser, 'Intention and Accident', *Philosophical Studies* 98 (2000), pp. 17–36. See discussion of The Big Bomb below.

'a threat (the second trolley) already present in his environment'.[33] But whether we count a threat as 'present' in someone's environment is simply a matter of something like how far it must move in order to kill him, and this clearly has no intrinsic moral significance. If we were sure it would kill Bugsy (and only Bugsy), it wouldn't matter if pushing the button in Start-Up caused the second trolley to travel from the south-west tip of Portugal to the north-east tip of Siberia before running him over.

To confirm the irrelevance of such factors, suppose you had three options: (1) push one button that sends a trolley into the missile and causes a second trolley to start moving and hit Freddy and Suzy, (2) push another button that causes the frontal bombs of a drone to fall on the missile and causes its rear bombs to fall on Bugsy's house, or (3) do neither 1 nor 2. Clearly it would be indefensible to choose option 1 over option 2. Something about its being the function of the second button to kill or its introducing a threat into Bugsy's environment is a ridiculous justification for killing an additional child or (if Suzy wasn't centred on the track) for exposing an additional child to a 1/11 risk of death or depriving her of seven years of life.

Our fifth step asks whether it would matter if the bombs you drop on Bugsy explode chemicals or kill him more 'directly', as in:

> *Dual-Action Bomber 2*. The only way to prevent a missile from killing five children is to press a button that will both drop your frontal bombs on the missile and drop your rear bombs on Bugsy's house. But this time Bugsy has no reinforced concrete or basement chemicals, and the rear bombs are duds, which will penetrate Bugsy's roof and crush him themselves.

Surely there is no moral difference between pressing the button in Dual-Action Bomber and Dual-Action Bomber 2. It seems obvious that it cannot matter whether your bombs kill Bugsy by exploding chemicals or crushing him.

Frances Kamm has, however, proposed a view according to which it would appear to be permissible to drop your bombs in Dual-Action Bomber but wrong to drop them in Dual-Action Bomber 2. According to Kamm, it is wrong to save five when 'something our means brings along with it causes [Bugsy's death] . . . *directly*', but permissible to save five when 'what our means bring along with them *indirectly* causes' his

---

[33] This might be suggested by Kamm, *Intricate Ethics*, p. 149. But Kamm seems to place relevance on the distinction between causes of harm that are 'introduced' as opposed to 'already present' only when they cause harm 'directly' (see discussion of Dual-Action Bomber 2 below).

death. She contrasts a case where to save five from dying you must drive them to the hospital, but driving to the hospital will eject lethal acid from a tank on your car onto Bugsy (Acid), with a case where driving the five to the hospital will produce vibrations that will cause rocks at the roadside to fall and crush Bugsy (Vibrations). Kamm claims that in Acid, something you need to save the five (your car's moving to the hospital) 'introduces something (acid) in a context where it directly harms someone', while in Vibrations what you need to save the five (your car's moving) merely causes things 'already present in the context' (rocks) to 'directly' cause death. In the same way, it would seem that in Dual-Action Bomber 2, something you need to save the five (flying and pushing the button where you do) 'introduces something' (the rear bombs) 'into a context where they directly harm' Bugsy, while in Dual-Action Bomber, what you need to save the five merely causes things 'already present in the context' (chemicals in the basement) to 'directly harm' him.[34]

As I mentioned, whether a cause of death is 'already present' or 'introduced into' a context is simply a matter of something like the distance it must travel to harm someone, and whether something causes harm 'directly' or 'indirectly' is simply a matter of something like whether it harmfully interacts with a victim's body or causes something else to do so. As I also mentioned, these factors seem as obviously devoid of intrinsic moral significance as skin colour. Kamm fails to make even a half-hearted attempt to shoulder the extreme burden of proof against the relevance of such manifestly irrelevant factors. After claiming that the important moral distinction is between 'subordinating' Bugsy to the five and merely 'substituting' Bugsy for them, Kamm admits that she can see no reason why there is any difference in terms of 'subordinating' and 'substituting' between using means that cause harm 'directly' as opposed to 'indirectly'. Indeed, she says 'the fact that we should not allow ourselves to do even the latter if a greater good were *not* at stake suggests that the weight of the greater good . . . is being transmitted so as to elevate the use of the means relative to the importance of not harming a person [thus *subordinating* this person]'.[35]

---

[34] Kamm, *Intricate Ethics*, p. 149. Note that because the reinforced concrete would otherwise protect Bugsy in Dual-Action Bomber, the basement chemicals are a necessary (not just overdetermining) 'direct' cause of Bugsy's death, and because the chemicals are absent in Dual-Action Bomber 2, the rear bombs are a necessary (not just overdetermining) 'direct' cause of his death (cf. n. 20).

[35] Kamm, *Intricate Ethics*, p. 166. It should be obvious, however, that this criterion of the 'transmission' of the importance of the five to our means fails to distinguish between *any* cases in which what we do will result in the survival of the five and the death of the one. Since 'we should not allow ourselves' to fail to *save* one individual at relatively trivial cost to ourselves 'if a greater good were not at stake', we can conclude that the

To confirm the irrelevance of whether your means of saving the five causes something that more 'directly' or more 'indirectly' kills Bugsy, suppose you had three options: (1) push one button that causes your frontal bombs to fall on the missile and causes your rockets to launch into Freddy and Suzy's house, which will cause vibrations that detonate chemicals in the protected basement in which they are hiding, (2) push another button that causes your frontal bombs to fall on the missile and causes your rear bombs to fall on and crush Bugsy, or (3) do neither 1 nor 2. It seems obviously indefensible to choose option 1 over option 2. That your bombs would kill Bugsy directly without the mediation of something like a chemical explosion is an unconscionable justification for killing an additional child or (if Suzy was partly shielded from the explosion) for exposing an additional child to a 1/11 risk of death or depriving her of seven years of life.

Our sixth step asks whether it would matter if the button released a single bomb that both destroys the missile and kills Bugsy, as in:

> *The Big Bomb*. The only way to prevent a missile from killing five children is to press a button that will drop a bomb on the missile. Unfortunately the bomb is so big that Bugsy in his nearby home will be enveloped in its blast.

It seems clear on reflection that there is no moral difference between pressing the button in Dual-Action Bomber 2 and The Big Bomb. Surely it cannot matter whether the button destroys Bugsy and the missile with different bombs or the same one.

One attempt to distinguish between Dual-Action Bomber 2 and The Big Bomb is suggested by Matthew Hanser. Hanser claims that 'you do *B* intentionally' if you act with the intention of doing *A* and you know that *A* will 'non-accidentally' cause *B*. According to Hanser, *A* 'non-accidentally' causes *B* if *A*'s happening is 'sufficient to explain' *B*'s happening, and he suggests that this is true if *A*'s happening 'formally explains' *B*, or *A*-type events have the function of causing *B*-type events. Hanser suggests that if you drop a nuclear bomb on a factory and it destroys neighbouring houses, the nuclear bomb has the function of destroying everything in its blast-radius, so its detonation suffices to explain the destruction of the houses, and you destroy them intentionally. On the other hand, he suggests that even if you can predict that dropping small conventional bombs on a factory will cause some of them to fall on neighbouring houses, they do not have

weight of the greater good of saving the five 'is being transmitted to elevate' our means of saving five 'relative to the importance of saving a person', thus 'subordinating' the one we fail to save to the five we do save.

the function of doing this, so your dropping them does not suffice to explain the destruction of the houses, and you do not destroy them intentionally.[36] Following Hanser, one might try saying that in dropping the big bomb you kill Bugsy intentionally, but in pressing the button that drops your rear and frontal bombs you do not.

But this would be a poor application of Hanser's analysis. There is just as much reason to say that pushings of the button that releases both the frontal and rear bombs have the function of destroying everything in the blast radii of both sets of bombs as there is to say that droppings of the big bomb have the function of destroying everything in its blast-radius. Just as engineers designed the mechanism for releasing the big bomb to deliver its blast to the point on which it lands, they designed the mechanism for releasing the frontal and rear bombs to deliver their blasts to the points on which they land. More importantly, whether you do something 'intentionally' in Hanser's sense cannot have intrinsic moral importance. On his analysis whether we do B intentionally depends on whether what we intend in doing B is 'sufficient to explain B', but whether it is depends on which parts of the full causal history of B we find salient.[37] Obviously, which parts of B's causal history we find salient is as devoid of intrinsic moral importance as skin colour. Moreover, Hanser's favoured generator of salience – facts about the 'functions' of artefacts – are historical facts about the intentions of the people who constructed them, which, as noted above, are obviously devoid of intrinsic moral importance.

A second attempt to distinguish between Dual-Action Bomber 2 and The Big Bomb is suggested by William Fitzpatrick. According to Fitzpatrick, if you intend to bring about state of affairs S, and S constitutes another state of affairs S', then you intend to bring about S'. Fitzpatrick claims that if you drop munitions that destroy an area containing both your target and civilians, your intended means of destroying the target include the entire area being destroyed, and this area being destroyed constitutes the destruction of the civilians, so you intend the death of the civilians. By contrast, he claims that if you drop munitions on a factory knowing that 'fallout' will kill nearby civilians, your intended means need not include the destruction of the area the civilians occupy, so you need not intend their deaths.[38]

[36] Hanser, 'Intention and Accident', esp. pp. 22–4, 26–7, and 31–2 (including n. 19).

[37] On the absence of a salience-independent metaphysical criterion of what (short of a full causal history) is 'sufficient to explain' an event – just as there is no salience-independent metaphysical criterion of what is 'the cause' of an event – see the various contributions to *Causation and Counterfactuals*, ed. J. Collins, N. Hall and L. A. Paul (Cambridge, MA, 2004).

[38] Fitzpatrick, 'The Intend/Foresee Distinction and the Problem of "Closeness"', *Philosophical Studies* 128 (2006), pp. 585–617, esp. 594–7 and 599–600.

Following Fitzpatrick, one might claim that in dropping the big bomb, your intended means of destroying the missile include the destruction of an area containing Bugsy, so you intend Bugsy's death, but in pushing the dual-purpose bomb release button you need not intend the destruction of the area around Bugsy, so you need not intend his death. Given a version of the Doctrine of Double Effect [DDE] according to which there are stronger intrinsic moral reasons against inflicting harms we intend than harms we merely foresee, this might make the first act wrong but the second permissible.

But this cannot be right. First, there is just as much reason to say that you intend the destruction of the area destroyed by your rear bombs as there is to say that you intend the destruction of the entire blast-radius of the big bomb. The destruction of the area destroyed by your rear bombs does nothing to destroy the missile, but this is equally true of the destruction of those parts of the big-bomb's blast radius that aren't actually occupied by the missile. Since in both cases the only thing that serves your purpose is the destruction of the area occupied by the missile, this is all you intend to destroy.[39] Second, while Fitzpatrick's criteria for when one state of affairs constitutes another are unclear, if a state of affairs you intend constitutes a distinct state of affairs, you need not intend the second state. If Fitzpatrick is right that your skull's being flattened constitutes your dying, then driving your car or wearing your clothes each day constitutes their being worn out. But as R. G. Frey observes, while we intend to wear our clothes and drive our cars each day, we merely foresee and do not intend their being worn out.[40]

Finally, Fitzpatrick is wrong to suggest that it matters intrinsically whether what we intend 'constitutes' or merely causes harm. Surely it is no more permissible to engage in area bombing if *you know*

---

[39] Fitzpatrick probably had in mind cases in which you destroy an entire area, not because your munitions cannot destroy one part without destroying another, but because you cannot tell where in the area your target is, so to be sure of destroying it you set out to destroy the entire area; cf. Thomas Nagel's description of area attacks during the Vietnam War (Nagel, 'War and Massacre', *Philosophy and Public Affairs* 1 (1972), pp. 123–44, at 130–1). In this sort of case, unlike The Big Bomb, destroying the entire area does serve your purpose of destroying your target.

[40] Frey, 'The Doctrine of Double Effect', *A Companion to Applied Ethics*, ed. R. G. Frey and C. Wellman (Oxford, 2003), pp. 464–74, at 467. I suspect that driving your car and wearing your clothes causes and does not constitute the damage in virtue of which they wear out, and that flattening your skull causes and certainly does not constitute the distinct states of your neurons dying or losing their capacity to support mental life. Still, I think you can intend one state of affairs and merely foresee another even when the first constitutes the second. Suppose I buy a six-pack of pop, and intend to drink one pop each day for the next six days. While my drinking a pop on each of the next six days appears to constitute my running out of pop, I think that I foresee running out, but do not intend to run out, of pop.

with certainty that, while the destruction of the intended area will not kill civilians outright, it will cause them to die of their wounds some seconds, minutes or days after you drop your bombs. Fitzpatrick might claim that the destruction of the area would still constitute the civilians' 'being seriously injured',[41] but there seems to be no moral difference between intending something that constitutes serious injury and intending something equally sure to cause death that doesn't constitute serious injury. Suppose we could save the five by destroying a missile in the vicinity of Bugsy's house by either (a) dropping the big bomb, which *we know with certainty* will kill Bugsy by causing him to die of his wounds in two days, or (b) bribing a cannibalistic giant to smash the missile by giving him both Freddy and Suzy, whom *we know with certainty* he will eat in two days. Should we really take option (b) because we would be intending a mere cause of death rather than a 'constituent' of serious injury?

   It is true that dropping the big bomb saves the five in virtue of its causing something (the bomb's explosion) that causes Bugsy's death rather 'directly', while the explosion that kills Bugsy in Dual-Action Bomber 2 does not help save the five. So in pressing the bomb-release in The Big Bomb you must intend something that kills Bugsy more 'directly' than anything you must intend in Dual-Action Bomber 2. But as we have seen, the mere number of causal links between what saves the five and what kills Bugsy is as obviously devoid of intrinsic moral

---

[41] Fitzpatrick's views on when a state of affairs constitutes someone's being 'seriously injured' are not spectacularly clear. He suggests that someone's being blown up or having her skull crushed constitutes her 'being seriously injured or killed', while he claims that in cutting an attacker in self-defence, her being sufficiently injured so as to be distracted does not constitute her being killed – or apparently even 'seriously injured' – although her 'death . . . may foreseeably follow later (perhaps because there is no medical aid available to stop the bleeding)' (604–7). Fitzpatrick suggests that his categorization is guided by whether the victim's dying is identical to 'or at least . . . nearly coincident' with some event that the agent believes causally necessary to her end. But whether what you need is 'nearly coincident with' the victim's death seems to amount to no more than whether her death occurs at roughly the same time as what you need, which is clearly of no intrinsic moral significance. Fitzpatrick acknowledges that there could be a spectrum of cases between crushing someone's skull and cutting her, and claims that there will be grey areas and matters of degree, but it is difficult to understand how he can suggest that these correspond to grey areas and differences of degree in our intrinsic moral reasons against inflicting harm. Permissive intuitions about killing in self-defence are misleading, since they may reflect the plausible idea that deontological constraints against inflicting harm do not apply to harming attackers. Apart from this, inflicting small cuts seems easier to justify because in real life they are much less likely to lead to death than crushing someone's skull or blowing someone up – there is always some (usually excellent) chance that blood loss or infection will not prove fatal. Absent these distractions, it seems obvious that there is no intrinsic moral difference between intending those states of affairs Fitzpatrick categorizes as constituting 'seriously injuries' and intending those he categorizes as constituting lesser injuries that you foresee with equal certainty will lead to death.

importance as skin colour. To confirm the irrelevance of this distinction, as well as those inspired by Hanser and Fitzpatrick, suppose you had three options: (1) push a button that will both cause your frontal bombs to fall on the missile and cause your rear bombs to fall on Freddy and Suzy, (2) drop the big bomb on the missile, the explosion of which will envelop Bugsy, or (3) do neither 1 nor 2. It seems manifestly indefensible to choose option 1 over option 2. You could not seriously try to justify your conduct by claiming that you merely 'killed Freddy and Suzy accidentally' or 'foresaw causing their lethal injuries' while pressing the second button would have 'been sufficient to explain Bugsy's death' or 'involved intending Bugsy's lethal injury'. This sounds false, and would in any event be an absurd justification for killing an additional child or (if Suzy was partly shielded) for exposing an additional child to a 1/11 risk of death or depriving her of seven years of life.

Our seventh step asks whether it would matter if you killed Bugsy in the course of eliminating a threat to the five by more 'indirect' means, as in:

> *Gettier Terror Bombing*. The only way to prevent a missile from killing five children is to demoralize the crew about to launch it into surrendering by convincing them that you have killed a child they care about. Your only way to do this is to explode fragmentation bombs in the air over a hospital in which the crew believes Timmy is trapped, which the crew will see from afar. Unbeknownst to the crew, Timmy has been safely evacuated. But also unbeknownst to the crew, Bugsy has been moved to the hospital and will be killed if you drop the bombs.[42]

In real life, Gettier terror bombing would be much less likely to save the five than dropping the big bomb. You could not be too sure that bombing the hospital would actually demoralize the crew into surrendering, as

---

[42] Dropping the bombs will cause the crew to believe that you have killed a child they care about in a way that resembles Edmund Gettier's examples of justified true beliefs that do not constitute knowledge (Gettier, 'Is Justified True Belief Knowledge?', *Analysis* 23 (1963), pp. 121–3). Cases like this are discussed by Mack, 'Three Ways to Kill', pp. 7–9; J. M. Fischer, M. Ravizza and D. Copp, 'Quinn on Double Effect: The Problem of "Closeness"', *Ethics* 103 (1993), pp. 707–25, at 720–2; Friedman, 'Minimizing Harm', p.169 n. 9; Fitzpatrick, 'The Intend/Foresee Distinction', pp. 612–13. Remarkably, Fitzpatrick suggests that on his account the Gettier terror bomber counts as intending Bugsy's death. But *clearly* he does not. All the Gettier terror bomber needs to save the five is the flash from the explosion some tens or hundreds of metres above Bugsy's head; the movements of bomb fragments towards Bugsy's location, let alone their wreaking havoc at that location, is nothing to his purpose. So all the Gettier terror bomber need intend is the exploding of the bombs tens or hundreds of metres above Bugsy's head, but this state of affairs – just like that of the trolley being diverted in Trolley – is *obviously* a mere cause and in no way 'constitutive' of Bugsy's dying or being injured.

opposed, say, to strengthening their resolve to launch the missile in revenge. But to determine whether there is any *intrinsic* difference between these kinds of bombing, suppose that in a cool hour the missile crew decided that if Timmy died, the costs of continued resistance would outweigh its expected benefits. Fearing that they might be moved to fight on by an unreasonable desire for revenge, they constructed a reliable device that would detect, via photo-sensor, any explosions over the hospital, and subsequently destroy the missile. If you knew all this, it seems obvious that there would be no moral difference between dropping the big bomb on the missile and exploding the fragmentation bombs over the hospital. Both acts cause explosions that destroy the missile and kill Bugsy; it's just that the second act destroys the missile by activating the photo-sensor.[43]

The fact that you drop the fragmentation bombs directly over Bugsy might make dropping them feel more like 'directly targeting' him than dropping the big bomb on the missile. But in both cases the destruction of Bugsy's actual location is nothing to your purpose. Dropping the big bomb saves the five by causing an explosion in the area occupied by the missile; that the explosion spreads horizontally to Bugsy is simply an unfortunate side-effect of creating it. In the exact same way, dropping the fragmentation bombs saves the five by causing an explosion over the hospital that causes the photo-sensors to destroy the missile; that the explosion spreads vertically to Bugsy is simply an unfortunate side-effect of creating it. Obviously, whether an explosion harms someone by spreading horizontally or vertically is as devoid of intrinsic moral relevance as skin colour.

To confirm the lack of intrinsic difference between doing something that kills Bugsy and 'directly' eliminates a threat to the five (dropping the big bomb) and doing something that kills Bugsy and 'indirectly' eliminates such a threat (Gettier terror bombing), suppose you had three options: (1) destroy the missile with the big bomb, knowing that both Freddy and Suzy are in its blast-radius, (2) destroy the missile

---

[43] One might argue that saving the five by demoralizing the crew is intrinsically harder to justify than saving them by activating the photo-sensor, because the first act involves 'terrorism' while the second does not. But I do not see how the mere fact of accomplishing one's goals by lowering the crew's morale could make a significant moral difference. The emotional harm they would suffer would be trivial in comparison to the lives of the five, and should presumably be discounted by the fact that they are culpable attackers. This said, it is not essential to my argument to consider cases that involve saving the five by demoralizing their attackers. I do so primarily to compare the causation of 'collateral damage' with terrorism. But my main points can be made by comparing the causation of 'collateral damage' with organ harvesting. So if you feel that there is a significant intrinsic difference between demoralizing the crew and activating the photo-sensor, feel free to reinterpret *all* my 'Terror Bombing' cases as involving mechanical devices (like photo-sensors) as opposed to demoralization.

by activating the photo-sensor by exploding fragmentation bombs over the hospital, knowing that Bugsy is inside, or (3) do neither 1 nor 2. The *bare* directness with which causes of Bugsy's death save the five – distinct from their having a greater chance of saving them – would be an unconscionable justification for killing an additional child or (if Suzy was partly shielded) for exposing an additional child to a 1/11 risk of death or depriving her of seven years of life.

Our eighth step asks whether it would matter if the crew was reliably tracking Bugsy, as in:

> *Lewissian Terror Bombing*. The only way to prevent a missile from killing five children is to demoralize the crew about to launch the missile into surrendering by convincing them that you have killed a child they care about. Your only way to do this is to explode fragmentation bombs in the air over a hospital in which the crew *knows* Bugsy is located (because they are reliably tracking his movements). The explosion will cause the crew, who will see it from afar, to surrender, but it will also kill Bugsy.[44]

It seems obvious to me that there is no moral difference between Gettier and Lewissian terror bombing. I cannot believe that it makes an intrinsic moral difference whether your killing Bugsy saves the five by getting the crew to believe falsely that you have killed Timmy or by getting them to believe truly that you have killed Bugsy.

However, Warren Quinn defended a version of the DDE according to which there is an important moral difference between Gettier and Lewissian terror bombing. Quinn observed that, because what causes the crew to surrender and saves the five is simply the flash of the explosion of the fragmentation bombs (rather than their fatal effects below), the Lewissian terror bomber need not intend any physical effects on Bugsy. But, Quinn noted, since the crew surrenders only because the bombs explode in Bugsy's vicinity, the Lewissian terror bomber 'does strictly aim at exploding bombs in [Bugsy's] vicinity . . . the bomber strictly intends to involve [Bugsy] in something (to make his bombs explode over [Bugsy]) in order to further his purpose precisely by way of [Bugsy] being involved'. Quinn's idea seemed to be that, because the crew is reliably tracking Bugsy, the five are saved only because Bugsy instantiates the property of *bombs exploding over him*, so the Lewissian terror bomber must intend that Bugsy instantiate this property. Moreover, Quinn suggested that this feature of Lewissian

---

[44] Warren Quinn reports the attribution of this example to David Lewis. See Quinn, 'Actions, Intentions, and Consequences: The Doctrine of Double Effect', *Philosophy and Public Affairs* 18 (1989), pp. 334–51, at 343 n. 16.

terror bombing is morally important. He urged us to accept a version of the DDE according to which it is harder to justify acting with the intention of causing a victim to instantiate a property in virtue of which she will be harmed than it is to act with the mere foresight that you will cause a victim to instantiate such a property.[45]

Quinn proposed a 'Kantian' rationale for his version of the DDE, suggesting that if an agent acts with the intention of causing victims to instantiate harmful properties, he treats them as means to his ends in a distinctively problematic way: 'He sees them as material to be strategically shaped or framed by his agency' and 'treat[s] them as if they were then and there for his purposes' (348). There are, however, deep problems with this rationale, many of which have been excellently articulated by Jonathan Bennett.[46] It is quite unclear what Quinn means by 'seeing someone as material to be shaped' or 'treating someone as if she existed for your purposes'. A Lewissian terror bomber who loves all the children equally and wishes only to save as many as possible certainly need not hold any silly metaphysical or teleological beliefs about 'what Bugsy is' or 'what Bugsy exists for'. The Lewissian terror bomber does intend to make a fact involving Bugsy (that bombs explode over him) obtain, which he knows, under these circumstances, will cause Bugsy's death.[47] But in what sense does this embody a more problematic attitude towards Bugsy than the Gettier terror bomber's intention to make a non-Bugsy-involving fact (that bombs explode over the hospital) obtain, which he equally well knows will kill Bugsy?

Quinn seemed closer to articulating a relevant difference between Lewissian and Gettier terror bombing when he suggested that a standard (or Lewissian) terror bomber's 'victims are not only harmed but, in some sense, used' (350). The Lewissian (unlike the Gettier) terror bomber does make use of Bugsy's presence, and the claim that 'You were just using me!' does seem to be a distinctive and plausible moral objection. But the injunction against 'using' others does not really support Quinn's version of the DDE. First, Quinn's actual formulation of the DDE makes it difficult to justify using, not *victims*, but (at most) *facts about* victims. For instance, Quinn's formulation makes it difficult to justify eliminating individuals who pose obstacles (like fatally blasting a fat man from a cave's mouth to save others from

---

[45] Quinn, 'Double Effect', pp. 341–4. Mack also seems to propose essentially this version of the DDE ('Three Ways to Kill', pp. 7–9).

[46] Bennett, *The Act Itself* (Oxford, 1995), pp. 218–21.

[47] It could very easily have been that this didn't cause Bugsy's death – the Lewissian terror bomber might have had evidence that, unbeknownst to the crew, Bugsy would be protected by a layer of reinforced concrete, but received last-minute intelligence that the concrete had actually been compromised.

drowning inside), but this does not use victims as means of saving others.[48]

Second, the plausible injunction against 'using' others is quite distinct from – and much less important than – the morality of inflicting and preventing serious harm. To complain plausibly that 'he was just using me' is to complain that you were wrongfully deceived by a manipulator into serving his purposes. The wrong is essentially one of dishonesty or infidelity to an implicit undertaking to tell the truth or provide important information.[49] We can justifiably complain of someone's being manipulated in this way even if it leaves her no worse off, and the complaint actually becomes infelicitous if the manipulation leads to serious harm. If someone tricks your friend into going on a date with him in order to make his wife jealous, and all this does is waste your friend's time and disappoint her expectations, it sounds right to complain that he 'used' her. But if he tricks your friend into

[48] Quinn was aware of this issue. After giving his formulation in terms of intended property instantiations (or 'harmful involvement'), he explicitly distinguished between actions that take advantage of the presence of a victim ('direct *opportunistic* agency') like Lewissian terror bombing, and actions that aim 'to remove an obstacle or difficulty that the victim presents (direct *eliminative* agency)' like blasting the fat man out of the cave, and held that the DDE 'strongly discriminates against' the first kind of actions and 'more weakly discriminates against' the second (344). (Incidentally, the only intuition Quinn used to motivate this was that it doesn't seem much harder to justify saving a mother's life by crushing her foetus's skull than fatally removing her foetus along with her uterus, which is complicated by issues of foetal moral status and the mother's right to control her body). Quinn conceded that 'it is less plausible . . . to think of the victims of direct eliminative agency as used', but insisted that 'Someone who gets in your way presents a strategic problem', and that this rationalized its being somewhat harder to justify direct eliminative agency than harming victims without intending them to instantiate any properties ('indirect agency').

But once we dismiss Quinn's obscure claims that to intend that someone instantiate a harmful property you must 'see her as material to be strategically shaped' or 'treat her as existing for your purposes', all we are left with are his remarks about victims being used. And these, it seems, do not support even a weaker form of discrimination against direct eliminative agency, since, as Bennett observes, it is not just less plausible, but entirely implausible to say that in eliminating victims who pose obstacles we use them as means (220). All these victims do is 'ennoble' or raise the status of your eliminating them to that of a cause of the survival of those you save, and as Stephen Yablo observes, things that simply make it necessary for other things to happen if an event is to happen are certainly not causes of that event or means of bringing it about (Yablo, 'Advertisement for a Sketch of an Outline of a Prototheory of Causation', *Causation and Counterfactuals*, ed. J. Collins, N. Hall and L. A. Paul (Cambridge, MA, 2004), pp. 119–37, at 120).

[49] You can implicitly undertake to tell the truth or provide information simply in virtue of conventional norms governing assertions, expressions of interest, or acceptances of offers. On the reduction of moral reasons not to lie to moral reasons to honour implicit undertakings, see W. D. Ross, *The Right and the Good* (Oxford, 1930), p. 21; David McNaughton, 'An Unconnected Heap of Duties?', *The Philosophical Quarterly* 46 (1996), pp. 433–47, at 437. For a similar account of the complaint that 'you were just using me!' and the wrongness of such using in terms of violations of obligations to provide information, see T. M. Scanlon, *Moral Dimensions: Permissibility, Meaning, Blame* (Cambridge, 2008), pp. 106–21.

going on a date in order to kill her – or even to non-fatally harvest one
of her kidneys to sell on the black market – it would be a bizarre moral
understatement to complain that he 'used' her. The manipulation would
not be worth mentioning given the much larger issue of his seriously
harming your friend.

Finally, to the extent that Bugsy is actually a subject of ethical
concern, the Lewissian terror bomber (who does not deceive or
manipulate Bugsy) need not intend that *Bugsy* instantiate any
properties. Suppose that Bugsy and his twin were in an accident
that destroyed Bugsy's body and his twins' cerebrum, and surgeons
transplanted Bugsy's cerebrum into his twins' body, resulting in an
individual who awakens with the same mental states as Bugsy. It seems
clear that this would be *just as good* for Bugsy as survival: he should
expect to have the experiences of the individual in his twin's body, and
those who care for Bugsy should, out of this care, strongly prefer the
transplant and want the individual in his twin's body to be as well-off
as possible. So to the extent that Bugsy is a subject of ethical concern,
he (who survives) is distinct from his body (which does not).[50] But the
Lewissian terror bomber need not intend instantiations of properties by
Bugsy, the subject of welfare who goes where his cerebrum goes. Since
the missile crew is only able to track Bugsy by tracking his body, they
would not notice if your agents moved Bugsy out of the hospital in a
glass jar before you dropped your bombs. Indeed, we can imagine that,
in Lewissian Terror Bombing, you had arranged for this to happen,

---

[50] See Sydney Shoemaker, *Self-Knowledge and Self-Identity* (Ithaca, 1963), p., 22; Eric
Olson, *The Human Animal: Personal Identity without Psychology* (Oxford, 1997), pp. 42–6.
Even Olson, who defends the view that we are organisms, does so by divorcing numerical
identity from what matters ethically, and suggests that we might use something like
the notion of 'the same person as Bugsy in the practical sense' to describe what I am
describing as 'Bugsy as a subject of ethical concern' (ch. 3, esp. pp. 65–71). On Olson's
view, we might have to admit that properties of Bugsy's body are properties of Bugsy, but
since our subject of ethical concern with Bugsy is not really Bugsy himself (an organism)
but Bugsy*, the welfare subject who inhabits Bugsy, they would not be properties of our
subject of ethical concern.

As Olson (following authors like Shoemaker and Derek Parfit) observes, cases of fission
give us good reason to think that the question of what matters in Bugsy's survival is not
the same as the question of what happens to entities to which Bugsy is numerically
identical. (See Shoemaker, 'Persons and their Pasts', *American Philosophical Quarterly*
7 (1970), pp. 269–85, at 284; Parfit, 'Personal Identity', *Philosophical Review* 80 (1971),
pp. 3–27. It seems as good as survival for Bugsy for half of his brain to be transplanted
into two bodies, resulting in two individuals with whom Bugsy is fully psychologically
continuous. But we cannot say that Bugsy is identical to one but not the other of these
individuals without being arbitrary, and we cannot say that he is identical to both
without the absurd consequence that they are identical to each other.) But I think that
Olson's total divorce of identity from what matters in survival goes too far. Following
Jeff McMahan, I think it best to tie the account of our numerical identity as closely as
possible to what matters in our survival – roughly, to a non-branching, minimal degree of
what matters in survival (McMahan, *The Ethics of Killing* (New York, 2002), pp. 51–5).

and for Bugsy to be safely transplanted into a back-up body that you have waiting for him. But, at the last minute, you are informed that the doctor was killed before he could extract Bugsy from his current body. Now it should be clear that, if you proceed with the raid, you must intend that Bugsy's current body (which the crew is tracking) instantiates the property of bombs exploding over it, but you merely foresee and do not intend that Bugsy, whose being trapped in that body is nothing to your purpose, instantiates the property of bombs exploding over him.

Should it really matter whether, in doing what you know will harm Bugsy, you intend that his body instantiates some property? Suppose your initial evidence was that you would be Gettier terror bombing – that the crew thought Timmy was in the hospital, but unbeknownst to them, Timmy had been evacuated and Bugsy had been taken to the hospital, so that the bombs exploding over Bugsy's body would play no role in saving the five (or a bazillion[51]) children. Suppose you were about to drop your bombs, when you received news that the crew has discovered that Timmy is gone and Bugsy is in the hospital. Should you really change your mind about whether to drop your bombs on the grounds that, because you now know that the location of Bugsy's body matters to the salvation of the five (or bazillion), you would be intending not only that the bombs explode over the hospital – which you already

---

[51] Quinn himself probably would have thought it wrong to save five by Gettier *or* Lewissian terror bombing Bugsy, since he defended a version of the Doctrine of Doing and Allowing (DDA) on which the greater difficulty of justifying doing harm in relation to allowing harm makes it wrong to kill one to save five (Quinn, 'Actions, Intentions, and Consequences: The Doctrine of Doing and Allowing', *Philosophical Review* 98 (1989), pp. 287–312). Presumably, he thought strategic bombings that kill innocents were permissible, while terror bombings with the same (expected) 'consequential profile' were wrong, only when the harms they could be expected to prevent were more than five times as great as the harms they could be expected to inflict. While my immediate aim is to argue against using Quinn's DDE to solve the trolley problem, my broader aim is to argue that there is *no* intrinsic difference between 'collateral damage' and directly inflicted harm. I thus consider making the five children into 'a bazillion' – a number such that it is morally permissible to kill one individual to save that number of individuals.

(Incidentally, Quinn believed – falsely – that his version of the DDA treated the choice of whether to divert in Trolley as a choice between killing Bugsy and killing five, and thus permitted diverting. On Quinn's version of the DDA, you count as doing harm if harm is caused by your actions *or* actions of objects or forces under your control that you deliberately fail to prevent. Quinn claimed that if someone *fails* to divert the trolley, 'it is because he intends the [trolley] to continue in a way that will save [Bugsy]. But then he intends that the [trolley] continue forward past the switch, and this leads to the death of the five' ('Doing and Allowing', p. 305). But in failing to divert the trolley to save Bugsy, you need only intend that it *not* continue in a way that threatens him. That it continues at all, and certainly that it continues past the switch, is nothing to your purpose of saving Bugsy – he would be saved equally well if the trolley simply stopped or fell off its track before reaching the switch. So in failing to divert the trolley to save Bugsy you need not intend that it continue, so Quinn's DDA counts you as merely *allowing* the five to die.)

knew would kill him – but intending that the bombs explode over his body? Such reasoning seems manifestly absurd.[52]

Or suppose that you had three options: (1) drop bombs over one hospital, in which the crew falsely believes Timmy and Lucy are trapped, while in reality they have been evacuated but unbeknownst to the crew Freddy and Suzy have become trapped in that hospital, (2) drop bombs over another hospital, in which Bugsy – whose body the crew is reliably tracking – is trapped, or (3) do neither 1 nor 2. On reflection it seems indefensible to choose option 1 over option 2. That killing Bugsy would have involved intending that bombs explode over his body, while killing Freddy and Suzy involved intending only that bombs explode over a hospital in which you knew them to be trapped, seems on its face to be an absurd justification for killing an additional child or (if Suzy was partly shielded) for exposing an additional child to a 1/11 risk of death or depriving her of seven years of life. Quinn's attempt to overcome this appearance of absurdity by claiming that killing Bugsy, unlike killing Freddy and Suzy, would have involved problematically 'using' someone fails because it mistakes a minor injunction against deceptively manipulating others for a major injunction against particular ways of harming others.[53]

---

[52] Only part of the absurdity stems from the fact that, as Thomson observed, the DDE as it is often understood implausibly tells us to 'look inward' and determine what our intention in doing something would be to decide whether to do it (Thomson, 'Self Defense', *Philosophy and Public Affairs* 20 (1991), pp. 283–310, at 293). Absurdity would remain if Quinn claimed that the explanatory relationship between harmful property instantiations and benefits itself matters intrinsically. Suppose you were about to engage in what you thought was Gettier terror bombing, when you received allegedly urgent news: 'Wait! Stop! The crew found out that Timmy is gone and Bugsy is in the hospital! If you explode the bombs, the explosion's being over Bugsy's body will play a role in explaining the salvation of the five (or bazillion)!' This reasoning also seems ridiculous.

[53] Mack proposed a 'Doctrine of Antecedent Causation', according to which 'When . . . inevitably injurious forces confront a person such that, no matter how that person acts, some nonaggressor(s) will be injured, the antecedent perilous forces bear the predominant causal responsibility for the subsequent injuries – unless those injuries are intended by a person whose action is an intermediate cause' ('Three Ways of Killing', p. 17). Although Mack speaks of 'injuries' being intended, his discussion of a Quinn-like way of getting the DDE to condemn acts like Lewissian terror bombing makes one suspect that he really means 'unless properties of nonaggressors in virtue of which they will be injured are intended by a person whose action is an intermediate cause'. Mack's suggestion would then be that if you kill Freddy and Suzy, the situation was primarily responsible for their deaths, but if you kill Bugsy, it is you rather than the situation that was primarily responsible for his death. But reading Mack's Doctrine this way makes it straightforwardly preposterous. There is no reason at all why (and it sounds self-evidently ludicrous to assert that), in killing someone by exploding bombs over him, you become primarily responsible for his death if you intend that they explode over his body, but your situation bears primary responsibility if you intend that they explode over the building in which he is trapped knowing full well that they are in fact exploding over his body.

Our ninth step asks whether it would matter if what saved the five was a physical effect on Bugsy's body, as in:

> *Standard Terror Bombing*. The only way to prevent a missile from killing five children is to demoralize the crew about to launch the missile into surrendering by convincing them that you have killed a child they care about. Your only way to do this is to drop a bomb on Bugsy, causing him to die, and causing the crew to see his lifeless body.

It seems obvious to me that there is no moral difference between Lewissian and standard terror bombing. I cannot believe that it makes an intrinsic moral difference whether your killing Bugsy saves the five by getting the crew to see the explosion that kills him or by getting them to see his body.

Since standard terror bombing saves the five only in virtue of the crew's seeing Bugsy dead, it might seem that the standard (unlike the Lewissian) terror bomber must actually intend Bugsy's death as a means of saving the five. If, as a traditional reading of the DDE claims, it is harder to justify inflicting harms that we intend than harms that we merely foresee, the standard terror bomber's intending Bugsy's death would make his act harder to justify than the Lewissian terror bomber's. But an argument of Bennett's shows that that the standard terror bomber need not intend Bugsy's death either. Bugsy's death consists in the permanent inoperativeness of his bodily or mental activity, but all the standard terror bomber needs to save the five is for Bugsy's body to remain inoperative long enough to be seen by the crew and cause them to surrender. Of course, under the circumstances this temporary inoperativeness cannot be caused without also causing permanent inoperativeness, but this seems no different from the Lewissian terror bomber's inability to cause an explosion over Bugsy's head without also causing fragments to rain down and kill him.[54]

---

[54] Bennett, 'Morality and Consequences', *The Tanner Lectures on Human Values*, vol. 2, ed. S. McMurrin (Salt Lake City, 1981), pp. 110–11; and *The Act Itself*, pp. 210–13). In his more recent discussion, Bennett suggests that, to avoid counter-intuitive results, we should count agents as intending as means not only those states of affairs that they believe are causally necessary to achieving their ends, but other states that 'we have not the faintest idea what it would be like to have the means' of separating from those needed to achieve their ends. Thus, while Bennett concedes the possibility of Bugsy's flying into small pieces without dying, he suggests that because we cannot easily imagine a technology for reassembling Bugsy, a standard terror bomber who must demoralize the crew by causing them to see Bugsy fly into small pieces should count as intending Bugsy's death as a means.

Of course, Bennett would be the first to note that how easy it is for us to imagine something lacks intrinsic moral significance, and there are variants on which Bennett's criterion allows the standard terror bomber to foresee rather than intend Bugsy's death

Still, because the standard terror bomber needs Bugsy's body to suffer damage that he knows will in fact prove fatal, it would seem that the standard (unlike the Lewissian) terror bomber must intend for Bugsy to suffer foreseeably fatal injuries. It might seem plausible to interpret the DDE as holding that it is harder to justify inflicting these kinds of injuries, or committing foreseeably lethal 'assault' or 'battery', with the intention than with the mere foresight of doing so, which would give us grounds for saying that standard terror bombing is harder to justify than Lewissian terror bombing.[55] There are, however, ways of varying the case so the lethal physical effect you intend on Bugsy does not seem to constitute an 'injury' or 'battering'. Suppose that the only way to demoralize the missile crew was for them to see you gently strap a bomb-vest on Bugsy and set the timer to explode sometime after they surrender, which *you know with certainty* will kill Bugsy (because it cannot be removed or defused without exploding). Here you need only intend that the timer be set and Bugsy have the vest strapped on him, which is not *itself* his being injured or battered in any intuitive sense. But much as we saw with the cannibalistic giant, it seems obvious that it makes no intrinsic moral difference whether in knowingly killing someone the lethal effects you intend to have on her body constitute or merely cause injuries or batteries. If you could demoralize the crew and save the five by (1) gently strapping vests on Freddy and Suzy which

(if he knows that the bombs will render Bugsy such that he could be resuscitated by the best contemporary medical technology, but that the crew lacks this technology). But I think it is a mistake to abandon the view that agents intend as means only those states of affairs they believe causally necessary to achieving their ends. If an agent cannot easily imagine two states coming apart, it is likely that he conceives of them as the same thing, believes them both necessary, and thus intends them both as means. But if an agent knows the states are distinct, I see nothing wrong with saying that she intends one but not the other. A standard terror bomber cannot truly deny that he intended a destructive or fatal effect on Bugsy's body, but he can truly say that he foresaw but did not intend his permanently eliminating Bugsy or causing Bugsy to never again experience goods (in which the harm of death consists). Of course it would be preposterous to use this as a justification or excuse for killing Bugsy, but that, I suspect, is because inflicting foreseen harm is just as difficult to justify as inflicting intended harm.

[55] See Joseph Shaw, 'Intentions and Trolleys', *Philosophical Quarterly* 56 (2006), pp. 63–83, at 69–71; Mikhail, 'Universal Moral Grammar', p. 145; Mikhail, *Elements of Moral Cognition* (New York, 2011), pp. 133–6, 148–52. In places Shaw and Mikhail seem to speak of an absolute prohibition of all intended assaults or batteries, or at least all 'serious' (as opposed to 'trivial') such assaults or batteries. But what makes an injury or battery *morally* 'serious' is not the dramaticality of the bodily damage it involves, but its probability of proving lethal or otherwise depriving the victim of intrinsic goods or subjecting her to intrinsic bads like pain. It is just as difficult to justify giving someone a small cut as blowing her body to bits if (as is almost never actually true) it is just as sure to kill her and she is just as sure to survive if you don't do it (cf. n. 41). Similarly, if you could anaesthetize Bugsy before blowing his neck-down body to bits and, as soon as this causes the missile crew to surrender, quickly reassemble his body so that he awakens moments later without noticing a thing, there could be no serious moral objection to doing so.

you *knew with certainty* would kill them or (2) bombing Bugsy, which you *knew with certainty* would kill him, it would be preposterous to take option 1 on the grounds that you merely intended to set the vests and strap them in, foreseeing but not intending that this would injure or batter them.

There is, however, a legal sense of 'battery' in which your causing something simply to touch or move someone's body (without her consent) is *itself* your battering her if this moving or touching in fact causes her to be harmed. John Mikhail considers a version of the DDE according to which it is harder to justify committing 'battery' in this sense with the intention than with the mere foresight of doing so, which would entail that standard terror bombing *and* deliberately strapping a lethal bomb-vest on Bugsy are harder to justify than Lewissian Terror Bombing.

While Mikhail's primary interests are in descriptive psychology, he asserts that this version of the DDE '[makes] no mention of what I take to be morally irrelevant information, such as race, gender, religious affiliation, and so forth' (259). But this claim seems to reflect an inadequate appreciation of what the principle is actually saying. 'Intending battery' may sound morally relevant if it calls to mind actions that are intended to harm others seriously, or affect their bodies in other ways that are more likely to harm them seriously than actions that risk collateral damage. But what Mikhail's DDE is actually saying is that it's intrinsically harder to justify deliberately causing an outcome in virtue of which you know someone will be harmed *if that outcome involves something's touching or moving her body*. It seems obvious, however, that it is morally irrelevant whether you deliberately cause something to touch or move someone or deliberately cause it to do something else to her (like hover around her) that you know with equal certainty will harm her in the exact same way.

Suppose you could save the five by setting the timer on and strapping a *hovering* bomb vest around Bugsy, which would not touch or move him but follow him by detecting his movements and altering its location. Clearly, it is no easier to justify this than it is to justify strapping a bomb vest on Bugsy that follows him by gently touching his body. Suppose you could save the five by pressing a button that would cause a robot to put a bomb vest around Bugsy, and your initial evidence was that it would be a hover-vest. Suppose you were about to push the button, when you received information that the vest is actually an ordinary one. Should you really change your mind about whether to push the button on the grounds that, because you now know that the vest will stay lethally close to Bugsy by touching his body instead of hovering a few millimetres above it, you would be intending not simply that it hover around him but that it touch him? To do so would be ludicrous; the

information on which Mikhail fixates is *at least* as obviously irrelevant as information about Bugsy's race, gender or religious affiliation.[56]

To confirm the lack of intrinsic difference between Lewissian and standard terror bombing, suppose you had three options: (1) explode fragmentation bombs, which the crew will see from afar, over Freddy and Suzy whom they are reliably tracking, (2) bomb Bugsy, whose lifeless body will be seen by the crew, or (3) do neither 1 nor 2. It seems indefensible to choose option 1 over option 2. That killing Bugsy would have involved intending that explosions impact his body, while killing Freddy and Suzy involved intending only that explosions occur so close to them that you knew with equal certainty they would be killed, is clearly an absurd justification for killing an additional child or (if Suzy was partly shielded) for exposing an additional child to a 1/11 risk of death or depriving her of seven years of life.

Our tenth step asks whether it would matter if saving the five required deliberately destroying Bugsy's organs, as in:

> *Organ Targeting*. A crew has set their missile to launch and kill five children. They have connected a mechanism that can disable the launch sequence to sensors that will be activated only if they detect the absence of vital organs in Bugsy's body. You are some distance off with a large-calibre rifle, and your only way to stop the missile from killing the five is to obliterate each of Bugsy's major organs by shooting them.

It seems obvious that there is no moral difference between lethally injuring Bugsy in Standard Terror Bombing and Organ Targeting. It cannot matter whether you must fatally immobilize Bugsy or fatally eliminate his organs.

It might seem that Bugsy has a special claim on his organs, and that obliterating them does not '[allow] a person to keep things to which he has a claim'.[57] But could this really make obliterating his organs any harder to justify than dropping bombs that you know with equal certainty will kill him by disrupting his vital processes? Doesn't Bugsy (or his representatives) have just as much of a claim on our allowing

---

[56] Again, the absurdity is not simply that Mikhail's DDE tell us to 'look inward' to determine what to do. Suppose you thought that the button would cause a hover-vest to be put on Bugsy, and you were about to push it, when Mikhail burst in with allegedly urgent information: 'Wait! Stop! The vest isn't a hover-vest but an ordinary one! If you press the button, you'll be saving the five in virtue of a vest touching Bugsy's body instead of hovering around it!' This too is *at least* as preposterous as treating race, gender or religious affiliation as morally relevant.

[57] Michael Tooley, *Abortion and Infanticide* (New York, 1983), p. 215, following Thomson, 'Killing, Letting Die, and the Trolley Problem', p. 213.

his vital processes to continue as he does on our leaving his vital organs in place, because doing otherwise would in both cases kill him?[58] To confirm the equal weightiness of these claims, suppose a second mechanism will disable the launch sequence if it detects the cessation of Freddy's and Suzy's bodily processes, giving you three options: (1) bomb Freddy and Suzy, which will lethally stop their bodily processes, (2) obliterate Bugsy's organs, or (3) do neither 1 nor 2. Clearly, it would be indefensible to choose option 1 over option 2. That killing Bugsy would have destroyed his organs, on which he has a special claim, while killing Freddy and Suzy simply interrupted their vital processes, would be a preposterous justification for killing an additional child or (if the sensor was connected only to Freddy, and Suzy was partly shielded) for exposing an additional child to a 1/11 risk of death or depriving her of seven years of life.

Our eleventh and final step asks whether it would matter if, instead of saving the five by destroying Bugsy's organs to disable the launch sequence in Organ Targeting, you could save the five by harvesting Bugsy's organs and transplanting them into the five in Transplant. It seems obvious that there is no moral difference between these acts. Surely it cannot matter whether you lethally cause Bugsy's organs to be absent from his body in order to prevent a missile from killing the five or in order to transplant the organs into the five to prevent their dying of organ failure.

   The fact that you save the five by transplanting into them the same organs you remove from Bugsy might give Transplant a more saliently 'redistributive' character, calling on the thought that you are wrongfully giving what belongs to Bugsy to someone else. Tooley suggests that removing Bugsy's organs in Transplant may be wrong because '[Bugsy] has more of a claim to his organs – simply because they are his – than the other five individuals' and 'the choice [in Transplant] is . . . between allowing a person to keep things to which he has a claim, and giving them to someone else who does not have a comparable claim to them' (215–16). But clearly it cannot matter whether, in benefiting others by depriving Bugsy of that to which he has a special claim, you give that very thing to others or do something else with it that is useful

---

[58] If we could destroy and replace Bugsy's organs with equally good ones without killing him, causing him any pain, inconveniencing him or going against his wishes, it seems clear to me that there would be no moral objection to doing so, just as I see no objection to taking a few hairs or skin cells from someone's body without his consent if it does not harm him or contradict his wishes. Surely any objection to doing this (even if it *does* contradict his wishes) would be *nowhere near* as weighty as the objections to seriously harming him. Even if those with strong views about 'self-ownership' disagree, they should agree that self-ownership gives one just as strong a claim against others' interfering with one's vital processes as against their destroying one's vital organs.

to them. Surely stealing someone's property is no more objectionable than wantonly destroying her property in exchange for a payment from a third party equal to the value of her property. Stealing from the poor to give their meagre belongings directly to the rich is no less wrong than destroying their belongings as a sacrifice to Poseidon, in exchange for which he rains down on the rich manna of equal value to those belongings.

To confirm the moral equivalence of Organ Targeting and Transplant, suppose you had three options: (1) obliterate the organs of Freddy and Suzy with a large-calibre rifle to please Poseidon, who will in exchange for this obliteration heal the five, (2) remove Bugsy's organs and transplant them into the five, or (3) do neither 1 nor 2. It seems indefensible to choose option 1 over option 2. That in killing Bugsy you would have given his organs to the five, while in killing Freddy and Suzy you merely destroyed their organs to save the five, would be an absurd justification for killing an additional child or (if Freddy's death and a risk or lesser harm to Suzy would be sufficient to please Poseidon) for exposing an additional child to a 1/11 risk of death or depriving her of seven years of life.

V. WHERE DO WE GO FROM HERE?

I have thus argued that it is morally irrelevant whether, in doing what saves five and kills Bugsy, you: (1) kill Bugsy with something that would not have 'directly' killed the five, (2) kill Bugsy with something that would not have 'indirectly' helped kill the five, (3) kill Bugsy with things that are initially stationary, (4) kill Bugsy with things designed to have destructive effects, (5) save the five with means that cause something 'directly' to kill Bugsy, (6) save the five in virtue (or with the intention) of causing something that itself 'directly' kills Bugsy, (7) save the five in virtue of something that kills Bugsy and 'indirectly' eliminates a threat to the five, (8) save the five in virtue of Bugsy's body instantiating some property in virtue of which he dies, (9) save the five in virtue of Bugsy's body being lethally touched or moved, (10) save the five in virtue of Bugsy's organs being lethally removed, or (11) save the five in virtue of Bugsy's organs being lethally removed and transplanted into the five.

None of these differences could make the difference between permissibly and impermissibly killing Bugsy to save five. Moreover, because none of these differences can justify even exposing an additional individual to a 1/11 risk of death or depriving her of seven out of seventy-seven years of life, they cannot make small moral differences that cumulatively make the difference between permissibly

and impermissibly killing one to save five.[59] Thus, because these differences are all that separate saving the five by diverting in Trolley and saving them by harvesting Bugsy's organs in Transplant, and they cannot make the difference between permissibly and impermissibly killing one to save five, it cannot be permissible to divert in Trolley but impermissible to harvest Bugsy's organs in Transplant.[60]

If this is right, then either the intuition that it's wrong to harvest Bugsy's organs or the intuition that it's permissible to divert the trolley is mistaken. Well, which is it? While a full resolution of this question is beyond the scope of this article, I wish to conclude by explaining what I take to be the important considerations that bear on answering it, and why I think they favour the view that it is our intuition that we may permissibly divert the trolley that has to go.

The first important consideration is that the Doctrine of Doing and Allowing (DDA), according to which it is intrinsically harder to justify inflicting harm on others than it is to justify failing to help others, seems directly plausible. As Philippa Foot put it, it seems that 'it takes more to justify an interference than to justify the withholding of goods or services', and as Bernard Williams put it, it seems that 'each of us is specially responsible for what he does', including the harms our actions produce.[61] As we saw in section III, the fact that an act would kill someone seems on its face to be a perfectly good justification for taking an alternative course of action that kills no one, even if it saves fewer individuals.

An absolutist version of the DDA, on which no amount of benefit could justify inflicting any amount of harm, seems (to me at least)

---

[59] Note that my argument for the irrelevance of factors (1)–(11) considers cases where the later factors in the list are present *in addition* to the earlier factors. As such, I have not been assuming that the moral relevance of the factors would have to be 'additive' in the sense that each factor's relevance would not be affected by the presence or absence of the others. I have given the later factors on the list ample opportunity to make non-additive moral differences that are present only when the earlier factors on the list are present (for the idea – and plausible examples – of non-additive moral factors, see Kagan, 'The Additive Fallacy', *Ethics* 99 (1988), pp. 5–31; Jonathan Dancy, *Ethics without Principles* (Oxford, 2004), ch. 3).

[60] While my case against the intrinsic relevance of factors (1)–(11) does not depend upon any empirical story of how we might be tricked into thinking that they are intrinsically relevant when they are not, I cannot refrain from observing that (a) all of these factors seem to increase the salience of the fact that you are harming Bugsy in relation to the fact that you are saving the five, and (b) most of them would in real-world cases decrease the probability of saving the five or increase the probability of killing Bugsy. It thus seems to me quite likely that the greater salience or real-world probability of killing in relation to saving is responsible for our greater tendencies to judge it wrong to kill in cases where more of the factors are present.

[61] Foot, 'Killing and Letting Die', p. 181; and Williams, 'A Critique of Utilitarianism', *Utilitarianism: For and Against*, ed. Bernard Williams and J. J. C. Smart (New York, 1973), pp. 77–150, at 99.

quite implausible. If you could prevent serious harm (like death or significant suffering) to someone by inflicting a relatively trivial harm (like being jostled or pricked by a pin) on someone else, it would seem absurd to refuse to prevent the harm on the grounds that it would involve harming someone else. Similarly, if we knew with certainty that billions of innocents would survive if and only if we killed one other innocent, it seems difficult to deny that it would be permissible to kill the one. But it does seem plausible that it is *a great deal* more difficult to justify inflicting harm than it is to justify failing to prevent harm. Perhaps the most plausible version of the DDA would be one according to which the infliction of harm can be justified only by the fact that it is needed to prevent an enormously, catastrophically or mind-bogglingly greater amount of harm. What exactly constitutes a mind-bogglingly greater amount of harm will be vague, but I think that vague comparisons like this are unavoidable in any plausible form of ethical thought.

If the direct plausibility of such a version of the DDA can survive further critical scrutiny,[62] we would seem to have the same kind of reasons to accept it as we have to accept other plausible principles, like the Principle of Beneficence, which states that there are moral reasons to promote the welfare of others. We would then have the very same kind of evidence for the view that moral reasons against inflicting harm are much stronger than moral reasons against allowing harm as we have for the view that there are moral reasons against allowing harm in the first place. But if we must accept such a version of the DDA along with the Principle of Beneficence, the conclusion seems inescapable that, because diverting the trolley inflicts the harm of death on Bugsy and does not prevent a mind-bogglingly greater amount of harm to others, it is impermissible to divert it.

Our intuitions about Trolley may rebel at this conclusion, but as we have seen, our intuitions about cases are unsustainable. Still, there seems to be a more principled source of discomfort with the conclusion that we may not divert the trolley. This is that, according to this application of the DDA, whether we are permitted to do what results

---

[62] For a serious attempt to show that, on closer inspection, the distinction between doing and allowing harm amounts to something that, like skin colour, seems obviously morally irrelevant (viz. a difference between there being few as opposed to many ways we could act that would result in the harm), see Bennett, 'Whatever the Consequences', *Analysis* 26 (1966), pp. 83–102; 'Morality and Consequences'. For a persuasive response to these arguments, see Daniel Dinello, 'On Killing and Letting Die', *Analysis* 31 (1971), pp. 83–6; Quinn, 'Doing and Allowing'. Whether the DDA can in fact survive critical scrutiny is a serious question on which I think the jury is still out. Below I address what I take to be one of the strongest objections to the DDA, but a full defence of the principle is beyond the scope of this article.

in the most individuals surviving depends on who happened to be threatened. If the trolley had been headed towards Bugsy rather than the five, we would have been permitted to do what would have resulted in five surviving by allowing it to hit Bugsy. But now, simply because the trolley happened to be headed towards the five instead of Bugsy, the DDA requires us to do what results in only one surviving by allowing it to hit the five. This sort of dependence of what we may do on where the trolley was originally headed can, in Alastair Norcross's words, seem to implausibly 'invest . . . the course of nature with moral significance', and can seem inconsistent with Thomson's observation that 'there is no principle of moral inertia . . . no prima facie duty to refrain from interfering with existing states of affairs just because they are existing states of affairs'.[63] This second consideration suggests that we must be permitted to minimize harm even if this involves causing harm to befall those who are not already threatened in cases like Trolley. But since as we have seen there is no relevant difference between Trolley and Transplant, this supports the conclusion that our intuition that we may not harvest Bugsy's organs is mistaken, and that we are permitted do that as well.

I think, however, that that this second consideration is mistaken, and that the DDA does not implausibly invest the course of nature with moral significance. The DDA does *not* hold that you should not divert in Trolley simply because the trolley will hit the five in the course of nature. The DDA permits *all sorts* of alterations of the course of nature, including most beneficent alterations that make others better off than they would have been. The only alterations it condemns are those that make others worse off, which it condemns, not because they are alterations of the course of nature or existing state of affairs, but because they are inflictions of harm. The DDA does not hold that which set of individuals is threatened by the trolley has *basic* moral significance; it merely holds that it has significance as a determinant of whether your act inflicts harm, or makes someone worse off than she would have been absent your interference.

Still, as Bennett has observed, it can seem unclear why our reasons to avoid acting in ways that result in someone's suffering harm should be strengthened by the mere fact that the harm would not have occurred absent our interference.[64] I confess that this does not seem so unclear to me; I find Williams's suggestion that we have special obligations not

---

[63] Norcross, 'Killing and Letting Die', *A Companion to Applied Ethics*, ed. R. G. Frey and C. Wellman (Oxford, 2003), pp. 451–63, at 455–7; Thomson, 'Killing, Letting Die, and the Trolley Problem', p. 209.

[64] Bennett, 'Negation and Abstention: Two Theories of Allowing', *Ethics* 104 (1993), pp. 75–96, at 95–6.

to make others worse off plausible. But I think there is an even more plausible way of understanding the DDA.

When you prevent harm to the five by inflicting harm on Bugsy, you benefit the five *at Bugsy's expense*, in that you benefit them by making things worse for Bugsy. This is very different from simply choosing to save five swimmers *rather than* one swimmer, where you do not benefit the five by making anything happen that makes the one worse off. I think that it is *highly* plausible that it is harder to justify benefiting some individuals at the expense of others than it is to justify simply benefiting some individuals rather than others. The idea here is to reject the Machiavellian dictum that 'the ends justify the means' by holding that, if doing something will benefit some individuals at the expense of others, the benefit loses (some of) its status as a good reason to do that thing. If this is right, it is typically harder to justify inflicting harm than it is to justify allowing harm, not because reasons against inflicting harm are in themselves stronger than reasons against allowing harm, but because the benefits that would justify inflicting harm would come at the expense of others, and this weakens their justifying force.[65]

It thus seems to me that the benefits of killing, even as a side-effect in cases like Trolley, come at the expense of those killed and cannot be justified by such benefits as the salvation of five others. McMahan has recently objected to this sort of extension of deontological constraints against inflicting harm to cases of foreseen but unintended harm on the grounds that, 'Since almost all contemporary warfare involves the killing of innocent bystanders in large numbers, this view entails pacifism as a matter of practical policy.'[66] Even if McMahan were right about the entailment, I am sceptical of the force of his objection. If the most plausible set of moral principles we can find entails that we

---

[65] There are instances of allowing harm that benefit some at the expense of (i.e. in virtue of ensuring the occurrence of events that harm) others, such as refraining from treating someone's disease so that you can learn how to treat others by observing her disease's fatal progression. I think that these acts are particularly difficult to justify for the same reason that inflicting harm is difficult to justify, and the DDA is best understood as asserting that doings of harm are harder to justify than allowings of harm that do *not* have this feature.

[66] McMahan, 'Intention, Permissibility, Terrorism, and War', *Philosophical Perspectives* 23 (2009), pp. 345–72, at 361. McMahan also objects to this view on the grounds that it is unstable because 'some of the objections to the relevance of intention seem also to challenge the asymmetry between doing and allowing'. But my argument here does not rely on the general objections to the relevance of intention that McMahan discusses; it relies only on the fact that on reflection it seems manifestly absurd to hold that the differences between paradigm cases of inflicting foreseen but unintended harm (diverting in Trolley) and paradigm cases of inflicting intended harm (harvesting in Transplant) could justify a moral difference between permissibly and impermissibly killing one to save five.

should be pacifists, I think we should be pacifists. The mere fact that plausible principles tell us to revise our intuitions about cases is not a good reason to reject those principles – else racists could justifiably conclude that skin colour *per se* must be morally relevant.

But McMahan's claim that deontological constraints against foreseeably inflicting serious harm would entail *de facto* pacifism seems to presuppose that the constraints would be absolute. As I indicated, however, the most plausible versions of these constraints hold that we are justified in inflicting serious harm on innocents if (but only if) it is the only way to prevent a catastrophically or mind-bogglingly greater amount of serious harm. In real life, neither the benefits nor the harms of military operations are certain, but the stronger the reasons we have to avoid doing something, the less willing we should be to risk doing it, and the greater the expectation of benefit we need to justify this risk.[67] Thus, under conditions of risk, a plausible non-absolutist version of the DDA will require that the possible benefits of our acts, weighted by their probabilities, be greater than the possible harms inflicted by our acts, weighted by their probabilities, by a mind-boggling margin.

I do not think that it is impossible for contemporary military operations to have this property. Carefully targeted strikes against genuine terrorists who really are about to attack large numbers of innocents are probably capable of having it. Given the murderous ambitions of the Nazis in the Second World War, daylight raids on factories and fighting in populated areas in critical battles probably had this property as well.[68]

---

[67] This is, I think, a general way of putting the insights of Frank Ramsey, John von Neumann and Oskar Morgenstern about the relationship between the preferences we (should) have for probability distributions over the various properties our acts might have and the preferences we (should) have for our acts having those properties (Ramsey, 'Truth and Probability' (1926), reprinted in Decision, *Probability, and Utility*, ed. P. Gardenfors and N. Sahlin (New York, 1988), pp. 19–47; and von Neumann and Morgenstern, *Theory of Games and Economic Behavior* (Princeton, 1944)). Decision theorists are usually interested in the probabilities of our acts resulting in various outcomes, but the idea easily generalizes to the probabilities of our acts having other properties, like inflicting harm.

[68] Indeed, some have argued that given the horrors that would have resulted from a Nazi victory, even the terror bombing of German cities by the British before the tide turned against the Nazis was justified on account of its having mind-bogglingly greater expected benefits than expected harms (see Michael Walzer, *Just and Unjust Wars* (New York, 1977), pp. 251–68; and John Rawls, '50 Years after Hiroshima', *Dissent* 42 (1995), pp. 323–7). Walzer rightly makes this conclusion contingent on the assumption, which he does not endorse, that 'unless the bombers were used in this way, the probability that Germany would eventually be defeated would be radically reduced'. Although I am no historian, I cannot see how it would have been reasonable for the British to think that terror bombing German cities would radically increase the probability of victory. Since this is probably the best candidate for an actual-world example of terror bombing that was likely to do mind-bogglingly more good than harm, I do not believe any actual instances

But it probably is not easy for contemporary military operations to have the property of being reasonably expected to prevent mind-bogglingly greater amounts of serious harm than they are likely to inflict. A strike that is certain to prevent a terrorist from killing one innocent but has a 10 per cent chance (or even perhaps a 1 per cent chance) of killing another innocent does not seem to have this property, any more than doing what is certain to kill one innocent to save ten (or even perhaps a hundred) others. If we can guarantee no less than a 10 per cent or 1 per cent chance of seriously harming civilians who are near a target we wish to strike,[69] we should presumably be extremely hesitant to strike it, and strike only if the consequences of not striking are expected to be much worse than the certain death of an innocent individual. If despite our best efforts to minimize collateral damage the expected amount of harm we will inflict on innocents by fighting a war cannot be expected to exceed by a mind-boggling margin the harm we will prevent, we should on this view not fight that war. This will probably rule out many wars, but as the example of the Second World War illustrates, it will not rule out all.

If I am correct that there is no intrinsic moral difference between killing innocents as a 'side-effect' and killing them by directly targeting them, the greater ease of justifying acts that risk collateral damage must lie entirely with their greater ratio of expected benefits to expected harms. Because terror bombing is almost never likely to prevent mind-bogglingly (or even marginally) greater harms than it inflicts, it makes sense to think of non-combatants as, for almost all practical purposes, morally immune from direct attack. Because operations that risk some collateral harm to civilians are far more likely to prevent mind-bogglingly greater harms, it makes sense to think of them as morally easier to justify. But we make a grave mistake if we categorize unintended harmful effects as *in themselves* easier to justify, and a graver one if this categorization leads us to accept collateral damage as justified by a 'proportional' good any time a salient military objective cannot be achieved without it. If there is, as I have argued, no intrinsic moral difference between directly attacking someone and harming her as a side-effect, then the expected benefits of acts that risk collaterally harming innocents are proportional to their expected harms only if a

---

of terror bombing have had it. But Walzer argues persuasively that acts which harmed civilians and *did* radically increase the probability of winning the Second World War, which some acts that risked collateral damage *did* do, were likely to do mind-bogglingly more good than harm.

[69] See McNeal, 'The U.S. Practice', esp. p. 18 n. 67.

direct attack on those innocents would be justified if (*per impossibile*) its expected benefits exceeded its expected harms by the same margin.[70]

hnye@ualberta.ca