



# Trolleys, Transplants and Inequality: An Egalitarian Proposal

Peter Baumann<sup>1</sup>

Received: 17 October 2019 / Accepted: 16 April 2020 / Published online: 17 August 2020  
© Springer Nature B.V. 2020

## Abstract

This paper deals with the core version of the Trolley Problem. In one case (Trolley) many people favor an act (throwing a switch) which will bring about the death of one person but save five other persons (rather than letting five die and one go on living). In another case (Transplant) most people would refuse to “sacrifice” one person in order to save five other lives (rather than letting five die and one go on living). Since the two cases seem similar in all relevant respects, we have to explain and justify the diverging verdicts. Since I don’t find current proposals of a solution convincing, I propose an alternative one according to which (Transplant)—but not (Trolley)—violates two forms of equality. I also test the underlying egalitarian principle against other cases. I argue that it offers a good explanation and also a normative justification of our intuitive verdicts about the cases.

## 1 The Problem

Here are two moral choice situations. First,

### TROLLEY

A bystander notices a runaway trolley (whose driver has fainted after brake failure) on the track ahead of which are five people who will not be able to get off the track in time. The track has a spur leading off to the right. The bystander can throw a switch thus turning the trolley off the main track and onto the one on the right. Unfortunately there is one person on the right-hand track who will also not be able to get off the track in time. The bystander can either turn the trolley, bringing about the death of the one; or he can refrain from turning the trolley, bringing about the death of the five.

(after Thomson 1976, 206, and 1985, 1397)

---

✉ Peter Baumann  
pbauman1@swarthmore.edu

<sup>1</sup> Department of Philosophy, Swarthmore College, 500 College Avenue, Swarthmore, PA 19081, USA

Second,

TRANSPLANT

A transplant surgeon has five patients who need new body parts to survive—one needs a heart, the others need, respectively, a liver, one lung, another lung, and a pair of kidneys—but all are of the same, relatively rare, blood-type. By chance, the surgeon learns that a healthy person with that very blood-type just walked in for a routine exam. He can take the healthy person's body parts, bring about his death, and implant them in his patients, saving them. Or he can refrain from taking the healthy person's body parts, letting his patients die. (after Thomson 1976, 206)<sup>1</sup>

It seems very plausible—and most people seem to agree—that in TROLLEY the bystander may or even ought to (see, e.g., Frowe 2018) throw the switch while in TRANSPLANT the surgeon ought not to “harvest” the organs of the one person.<sup>2</sup> However, in both cases the choice is between bringing about the death of one person or the death of five persons.<sup>3</sup> In this respect the cases don't differ at all. If the diverging though equally plausible verdicts can each be defended as justified, then there must be some morally relevant difference between the cases. The problem of finding such relevant differences has proven very difficult; it is known as the “Trolley-problem”.<sup>4</sup> Insofar as we are facing a triad of mutually inconsistent but equally plausible claims (the verdict in TROLLEY, the verdict in TRANSPLANT, and the claim that the two cases don't differ morally), we are facing a paradox. So far, no single proposed solution has convinced a significant majority of authors.<sup>5</sup> In the following 2 sections

<sup>1</sup> Both case descriptions are simplified; each of them invites certain *ceteris-paribus* assumptions not spelled out explicitly. But this should not be problematic. However, if one still has doubts about the realistic character of the above two scenarios then one can consider decisions in war like Churchill's alleged decision concerning the “Coventry Blitz” (disputed historically but certainly still a realistic case), contemporary cases of illegal organ trade, certain mining accidents or shipwrecks, or the case of the fourth plane in the 9/11 attacks (there are even court decisions concerning similar cases). Wood (2011), Fried (2012) and Hare (1981, 139), however, argue (not convincingly in my view) that the consideration of Trolley-type cases is useless and at best very misleading. See also Bauman et al. (2014).

<sup>2</sup> What seems to be clearly out of the question is to turn a trolley away from one to five others (*ceteris paribus*) or to harvest five persons for their organs in order to benefit one person (*ceteris paribus*). This kind of “turn” seems clearly worse than the one presented in the above two cases.

<sup>3</sup> I am using the term “bringing about” as neutral between acts and omissions and between killing and letting die.

<sup>4</sup> The current discussion has been initiated by Foot (1978). Fischer and Ravizza (1992b, 1–15), Bruers and Braeckman (2014), and Edmonds (2014) offer overviews over the problem's discussion; also see Cathcart (2013). See also Ishiguro (2005) for a novel dealing with closely related topics.—There has been a lot of empirical (psychological, neuro-scientific, etc.) research about our reactions to trolley cases. See, e.g., Cushman et al. (2006), Di Nucci (2013), Greene et al. (2001), Hauser et al. (2007), Lanteri et al. (2008), Liao et al. (2012), Mikhail (2011), Navarrete et al. (2012), Petrinovich and O'Neill (1996), Rai and Holyoak (2010), Terbeck et al. (2013), Waldmann and Dieterich (2007), and Wiech et al. (2013). I won't go into this here.

<sup>5</sup> Foot 1978 invokes the difference between the more weighty negative duty not to kill and the less weighty positive duty to help. Thomson (1976, 1985) argues that in TRANSPLANT but not in TROLLEY bringing about the death of the one involves rights violations and the creation of a new threat (for a later position based on a view of what's to a person's advantage see Thomson 1990, ch.7; finally, see Thomson (2008, 2016) for the view that there isn't a problem in the first place because we're not allowed

I introduce my own proposal. Section 4 tests it against cases. Section 5 deals with borderline cases and Sect. 6 with *ceteris paribus* clauses. Section 7 concludes with an outlook on further questions and topics.

## 2 Two Kinds of Equality

I find myself unconvinced by any of these proposals but won't go into the reasons for my skepticism about them.<sup>6</sup> I rather intend to make a fresh start and propose and defend a new solution. I will propose as crucial a combination of two differences between cases like TROLLEY and TRANSPLANT that haven't been discussed so far. These differences have to do with two forms of inequality. They can help explain why we judge cases like TROLLEY and TRANSPLANT differently and also justify such diverging verdicts. I will aim at a reflective equilibrium between plausible theoretical principles (of equality here) and "intuitions" about cases (assuming the cases are not too far-fetched and complicated).

First, there is what one can call "exchangeability" in TROLLEY: If the person on the right track had been on the main track and one of the five persons on the main track on the right track instead, nothing relevant to the choice situation would have changed. There are no differences between the individual people on the two tracks that matter to the choice situation. There is, interestingly, no such exchangeability in TRANSPLANT: If one keeps the choice situation fixed, then the one healthy person isn't and couldn't (as a healthy person) be one of the five patients with organ failure (and none of the five original patients could play the role of the healthy person). The six people in TROLLEY are equal insofar as each could play the role of the one; the six people in TRANSPLANT are not equal in that respect.

---

Footnote 5 (continued)

to turn the switch in TROLLEY). Kamm (1989, 2000), Kamm (2001, chs. 6 and 7, 2016a, b) locates the morally relevant difference between the cases in the different causal profiles of the different acts. Costa (1986, 1987) applies the doctrine of double effect in order to locate a relevant difference between the two acts harming the one; for another defense of the doctrine see Kaufman 2016. For the presumed difference between killing and letting die (related to the difference between acts and omissions) see, amongst others, Locke (1982) or Norcross (2008, 68–72). For further proposed relevant differences see: Clark (1995), Gert (1993), Griffin (1996, 98–102), Hallborg (1997), Hanna (1992), Haslett (2011), Laycock (1985/86), Locke (1982, 474–475), Montmarquet (1982), Naylor (1988), Richardson (2008), Russell (1977), Schmidtz (2006), Shaw (2006), Sheng (1995), and Stark (2017).—One could make the following objection against my claim that there is no widespread agreement on the proper solution of the Trolley-problem: One could claim that there is, for instance, widespread agreement amongst consequentialists (e.g., "Use considerations of maximizing expected utilities and disregard recalcitrant intuitions") or amongst deontologists (e.g., "Don't use people as mere means, like in TRANSPLANT (though not in TROLLEY)"). However, none of these and other views constitute majority views; apart from that, there are very different consequentialist (permissible killing in both cases vs. impermissibility in TRANSPLANT because of the overwhelmingly bad consequence of missing trust in hospitals and doctors) as well as very different deontologist solutions of the Trolley-problem (Kant's means-end formula vs. the doctrine of double effect). (Also see the examples mentioned above). Thanks to a referee who pressed me on this issue!

<sup>6</sup> Going more into any of them would change the focus of this paper, make it much longer and still not change much at all about the overall argument to come.

Here is another way to describe the relevant difference between TROLLEY and TRANSPLANT. In TRANSPLANT the one person which is contrasted with five others has a property relevant for the decision to be made—being healthy—which predates the scenario and which the person possesses no matter whether she enters the scenario or not (passes by the hospital or not). The same holds for the other five persons: They are not in a good state of health, whether they enter this hospital or not; their state of health predates the scenario. And the one and the five differ with respect to the relevant property (healthy vs. not healthy). TROLLEY is different in this respect: The property relevant to the decision to be made—which track one is on—does not predate the scenario and is determined by the way the scenario is constructed; being located on this or that track is not a property any of the individuals possesses independently from the scenario. This is why it does not matter who is on what track in TROLLEY but why it does matter who (amongst those with pre-existing properties) is passing by the hospital in TRANSPLANT. People are thus “exchangeable” in TROLLEY but not in TRANSPLANT.<sup>7</sup>

Second, there are different profiles of gains and losses in TROLLEY and in TRANSPLANT. Consider the situation for the two involved “parties” both just before the making of the decision and just after the decision has led to the expected result.<sup>8</sup> If the switch is being thrown in Trolley, then the situation of the one worsens (considerably) while the situation of the five remains the same (see fn.8); if the switch is not being thrown, then the situation of the one remains the same while the situation of the five worsens (considerably).<sup>9</sup> We can capture this in the following way (with the changes for the act harming the one (throwing the switch) listed before the

<sup>7</sup> Thanks to a referee whose suggestions convinced me to rethink and rewrite a former version of this passage.

<sup>8</sup> We shall focus here on the more “direct” threats (e.g., to the five at the outset of TROLLEY) and put the more “indirect” threats (e.g., to the one at the outset of TROLLEY) aside as comparatively minor. By the way, at least TROLLEY could be easily changed in such a way that both sides are equally under threat, both directly and indirectly.—More importantly, we can leave the comparatively small gains or losses resulting just from the posing or removing of threats (getting anxious, feeling relieved) out of the picture here for the sake of simplicity. This doesn’t change the overall situation: Being alive or not matters much more than whether one is under some threat.

<sup>9</sup> But doesn’t throwing the switch improve the situation of the five (considerably) because otherwise they would have died? And doesn’t not throwing the switch keep their situation the same? This way of judging gains and benefits seems to presuppose that there is a “normal course of events” with which the outcome of the act of the person at the switch is being compared so that one can determine “relative” gains and losses. I have doubts about the applicability of the idea of a normal course of events in general but also in this case in particular: What determines normality when the choice situation leaves it (wide) open what will happen? One might propose to use a distinction between acts and omissions here and let the result of omissions determine the normal course of events. This proposal would be very much in need of argument, also given how controversial and problematic this distinction is; I doubt very much that a convincing case like that can be made. However, I won’t go into this here, also because I don’t need to. My comparison point for the determination of gains and losses is different: It is what results from the agent’s decision (to throw or not to throw); this is being compared with the status quo ante the decision (see above). This is a different way of determining gains and losses but the one relevant and used here. Thanks to Bruce Russell who raised this issue.

changes for the act harming the five (not throwing the switch);<sup>10</sup> and for both acts the changes for the one listed before the changes for the five):

#### Changes in TROLLEY

Worse-same/same-worse

Interestingly, the profile of gains and losses is different in TRANSPLANT. If the one is being “harvested” for his organs, then the situation for the one worsens (considerably) while the situation for the five improves with the transplant (considerably: they go from being ill to being well<sup>11</sup>); if the one is left unharmed, then their situation remains the same as does the situation of the five (alternatively one could argue that the five get worse because they get closer to death and will die soon). Hence, we get the following profile (again, with the changes for the act harming the one (“harvesting”) listed first and for both acts the changes for the one listed before the changes for the five):

#### Changes in TRANSPLANT

Worse-better/same–same (alternatively: worse)

There is thus an equality and symmetry in the gains-losses-profile for TROLLEY which is clearly missing in TRANSPLANT’s profile. In TRANSPLANT the one can end up much worse and at best remain unchanged while the five can end up better and at worst remain unchanged (by the decision, at least). The one cannot get better while the five can get better. The “cards” of the five are much better than the ones of the one.<sup>12</sup> We reject harming the one because the unequal distribution of potential gains and benefits seems to us to be unfair to him and unjust. The injustice of the choice situation prohibits exploiting it and harming the potentially disadvantaged one.

### **3 A Principle**

One can sum up the two above differences between our two cases by saying that in TROLLEY there is equality both with respect to the individuals’ suitability for playing any of the roles in the scenario (“role-equality”; “exchangeability”) and with respect to gains-losses-profiles (“risk-equality”) while in TRANSPLANT there is inequality in both respects.<sup>13</sup> This double difference explains, I propose, our diverging verdicts in both cases. And insofar as equality is a positive value here, it also justifies our different verdicts.

<sup>10</sup> I am using “to harm” as neutral between acts and omissions here, like “to bring about” (see fn.3 above).

<sup>11</sup> There is certainly a counterfactual element in the evaluation of the possible states of the subjects here. But this does, of course, not entail that counterfactual considerations play a role when comparing states to judge the equality in expected gains and losses (see fn.9 above).

<sup>12</sup> 72 of the 81 possible combinations of gaining, losing or remaining the same for two parties are cases of unequal gains-losses-profiles. It is not necessary to discuss each of these cases.

<sup>13</sup> I am using “role-equality” and “risk-equality” for lack of better terms.

We can generalize this idea in the following way:

#### Principle of Equal Harm (PEH)

In situations where one can either do harm (of a certain type and amount) to a smaller number of people (and prevent similar harm to a bigger number of people) or can do the same kind of harm to a bigger number of people (and prevent similar harm to the smaller number of people), harming the smaller number is

- permissible<sup>14</sup> if (and only if) there is both role-equality and risk-equality, but
- not permissible if (and only if) there is neither role-equality nor risk-equality.

NB: If (and only if) there is one form of equality but not the other, then the case is a borderline case.<sup>15</sup>

This principle is defeasible; it doesn't hold when certain *ceteris paribus* assumptions are not true (see Sect. 6 below). Apart from that, one could argue that (PEH) only holds for serious harms, like death. I would be fine with a restriction to serious harms; perhaps the gravity of harm makes a relevant moral difference here (a topic I cannot discuss here but see SAMARITAN below). Things will become considerably more complicated if the harm distributed on the two sides is of a different type and severity; I also cannot go into such topics here. The upshot of all this is that, given certain *ceteris paribus* assumptions, role-equality and risk-equality are both individually necessary and jointly sufficient for permissibility of harming the smaller rather than the greater number of people.<sup>16</sup>

## 4 Support from Cases

One can find confirmation for (PEH) in other cases. Consider, for instance, Bernard Williams' well-known Jim-Pedro example:

#### JIM-PEDRO

Jim finds himself in the central square of a small South American town. Tied up against the wall are a row of twenty Indians, most terrified, a few defiant, in front of them several armed men in uniform. A heavy man in a sweat-stained

<sup>14</sup> One might want to add "obligatory" to "permissible". I will leave this aspect aside here.

<sup>15</sup> These are cases of indeterminacy: it is neither (true that it is) permissible nor (true that it is) impermissible to harm the smaller number.—Sometimes, risk-inequality is being attributed greater weight than role-equality, or *vc.vs.* I have to leave the question open here why and under what conditions one aspect of (in-)equality might be weighed more than the other. I also have to leave the question open whether this is just a psychological bias or rather something that can be justified with good reasons (see also Sect. 5 on borderline cases and the possibility that our verdicts are quite unambiguous).

<sup>16</sup> (PEH) doesn't entail that "numbers count" in a decisive way (see, e.g., Taurek 1977); it rather only tells us that given certain conditions it is permissible to do harm to the smaller number of people and prevent harm to the greater number of people without claiming (nor ruling out) that this is because of the numerical difference as such.—I won't go into possible principles for distributing benefits (rather than harms) or into the question under what conditions it might be permissible to do lesser harm to one person and save another person from more harm.

khaki shirt turns out to be the captain in charge and, after a good deal of questioning of Jim which establishes that he got there by accident while on a botanical expedition, explains that the Indians are a random group of the inhabitants who, after recent acts of protest against the government, are just about to be killed to remind other possible protestors of the advantages of not protesting. However, since Jim is an honored visitor from another land, the captain is happy to offer him a guest's privilege of killing on of the Indians himself. If Jim accepts, then as a special mark of the occasion, the other Indians will be let off. Of course, if Jim refuses, then there is no special occasion, and Pedro here will do what he was about to do when Jim arrived, and kill them all. Jim, with some desperate recollection of schoolboy fiction, wonders whether if he got hold of a gun, he could hold the captain, Pedro and the rest of the soldiers to threat, but it is quite clear from the set-up that nothing of that kind is going to work: any attempt at that sort of thing will mean that all the Indians will be killed, and himself. The men against the wall, and the other villagers, understand the situation, and are obviously begging him to accept. What should he do?

(Williams 1973, 98–99)

It is plausible to suggest (and Williams would do so, too) that it is at least permissible for the agent (Jim) in this case to act in such a way that harm is done to the one and not to all of the twenty. In order to modify the case a bit so that it fits the structure of Trolley cases more closely, we can assume that Jim's potential unique victim is a 21st person from the area so that either the 21st or the other 20 will die. Here, (PEH) suggests again, that killing the one is at least permissible—as there is both role-equality here and risk-equality. Now, contrast Williams' JIM-PEDRO with the following case:

#### SENTENCE OR SACRIFICE

Five prisoners are on death row to be executed soon (unjustly, we may assume, and as "enemies of the State"). However, you know that if you hand over an unrelated, "non-enemy" person to the emperor he would throw that person to the lions in the circus and free the five prisoners. What should you do?

In this case it is quite plausible that one should not do harm to the one in order to save the five from it. (PEH) suggests this, too—as there is neither role-equality nor risk-equality.

Let us look at the following, potential counterexample, the case of the "Toxic Gas Transplant":<sup>17</sup> "Five people are trapped in room A. Before they can be liberated, they will become exposed to a toxic gas that destroys their lungs. The only way to save them, is by killing one person in another room B. If this person is killed

<sup>17</sup> Thanks to a referee!

right now, five pieces of his lungs can be used in time to save the five people in A. There are no other lungs for transplantation available, the lungs of the B person can be used in time if the B person is killed before the people in A are exposed to the gas.” The objector claims that it is impermissible to kill B even though there is both role- and risk-equality. Leaving out some details of the objection, I can say that I am not convinced that there is risk-equality in the situation as described. The people in A get better if B is being killed because they will be liberated (in contrast to being trapped) whereas B can at best remain the same. This is a crucial asymmetry.

However, the objector might (or should even) respond that being trapped in a room or being liberated is just a minor detail in this case. So, what if the people in A are not trapped in a room? Then we get a different case. In this case we do indeed have risk-equality, in addition to role-equality. But is it still impermissible to harm the one, as the objector would or should probably claim? To answer this question, consider a small change in the scenario: The agent has to choose where the gas will go—into the room with the five or into the room with the one. Now the scenario is very close to TROLLEY, and the intuition of impermissibility disappears or is at least very much weakened. However, since the specific way of bringing about the death of the one is not morally relevant, we should judge the objector’s modified case the same way we judge Trolley.

As a matter of psychological fact, it seems true that we’re typically feeling much more reluctant do bring about the death of the one in the original way rather than the second way. But as long as there is no moral relevance to the difference between the two ways of bringing death, we should rather interpret the different reaction as a merely psychological effect. The objector might want to identify a further morally relevant factor, independent from risk-inequality. But as long as no such further morally relevant factor is being proposed, the permissibility intuition (in the modified scenario) remains strong. Finally, couldn’t the objector turn the tables and argue from the impermissibility of harming the one in the original (objector’s) scenario to impermissibility of harming the one in the modified scenario as well as in TROLLEY? Isn’t one person’s *modus ponens* another person’s *modus tollens*? This would, indeed dissolve the initial problem and paradox. However, I take the claim of permissibility to throw the switch in TROLLEY to be so strong—and the claim that there is an initial Trolley-problem or paradox to be so convincing—that it is much more convincing to stick with the above response to the objection.

Finally, here is a pair of cases which one might think create problems for my account.<sup>18</sup> First, there is STRATEGIC BOMBER which is a case where a bomber pilot during war faces the choice of (1) dropping a bomb on a munitions factory to contribute to an earlier ending of the war while killing some civilians in the area or of (2) not dropping the bomb, preventing the deaths of the civilians but not contributing to a shortening of the war. Second, there is TERROR BOMBER which is a case where a bomber pilot during war faces the choice of (1) dropping a bomb on the

<sup>18</sup> Thanks to a referee who pressed on on this pair!



same number of civilians, thus demoralizing the enemy and shortening the war or (2) not dropping the bomb on the civilians but also not shortening the war (see, e.g., Stark 2017 on this and with respect to the Trolley-problem). In both cases there is role-inequality with risk-equality but we judge them differently morally. Let us assume for the sake of the objection that different people would die, depending on which decision the pilot makes. Let us also leave aside complications that might arise according to the doctrine of double effect because in the first case the deaths of the civilians would not be intended but only foreseen as a bad (side-)effect of one's action. It seems to me that this is a case where an additional factor comes into play: the immorality of using violence in order to terrorize and demoralize. For this see Sect. 6 below on *ceteris paribus* clauses. I don't want to claim that this is the only factor that makes a difference for moral evaluation but it I do think that it does make such a difference here.

## 5 Borderlines

For borderline cases, consider, for instance, the following two related cases. First,

### FOOTBRIDGE

George is on a footbridge over the trolley tracks. He knows trolleys, and can see that the one approaching the bridge is out of control. On the track back of the bridge there are five people; the banks are so steep that they will not be able to get off the track in time. George knows that the only way to stop an out-of-control trolley is to drop a very heavy weight into its path. But the only available, sufficiently heavy weight is a fat man, also watching the trolley from the footbridge. George can shove the fat man onto the track in the path of the trolley, killing the fat man; or he can refrain from doing this, letting the five die.

(Thomson 1976, 207–208)

Second (assuming Joe to be overweight),

### LOOP

The trolley is directed away from the five onto a track which, however, loops back toward them. The only reason it does not kill the five anyway, is that it slams into Joe who is seated on that track and whose weight stops it from going all the way around.

(Kamm 1989, 256, fn.6; see also Thomson 1985, 1402)

In both cases there is no role-equality<sup>19</sup> but arguable risk-equality (like in TROLLEY). If (PEH) is true and no other principle in force here, then this mix of equality and inequality explains and justifies reluctance to judge that harming the one is permissible. On the other hand, these two cases also don't strike one as ones where it is plausibly or even clearly impermissible to harm the one. They appear to be borderline cases. To be sure, many would be more inclined in LOOP than in FOOTBRIDGE to judge that harming the one is permissible. As mentioned above (see fn.15), different weights might sometimes be given to the aspects of risk and role, and perhaps even with good reason. However, I myself cannot see any relevant differences between these two cases: Whether one pushes the man or pushes a button which opens a trap door so that the man falls onto the track doesn't make a morally relevant difference; furthermore, whether one does the latter or turns a "Lazy Susan-type device" (Kamm 1989, 228–229) so that the five are not on track anymore but the one is now on the track also doesn't make a morally relevant difference; finally, whether one does the latter or throws the switch in LOOP also doesn't make a morally relevant difference; hence, FOOTBRIDGE and LOOP are morally on a par.<sup>20</sup> Given (PEH), all this suggests that both are borderline cases where we are justified in hesitating judging one way or the other.<sup>21</sup>

<sup>19</sup> What if the five are also overweight? Well, then the trolley would only hit and kill the first of them. In this case we would have a very different choice situation: an "indifference dilemma" like in drowning twin cases. A referee objected to this (thanks for this challenge!), presenting the following case: "Suppose for example that behind the bridge there is a downhill slope, such that if the trolley is not blocked under the bridge, it will go downhill and gain enough speed to kill five overweight people. Or suppose it requires five heavy items to block the trolley, and on the bridge there are four heavy stones and one overweight person that one could easily push off the bridge." The critical suggestion is that there is role-equality in this case, in addition to risk-equality but still no reason at all to think it is permissible to push the person from the bridge. I agree that there is role-equality here. Shouldn't I then conclude that it is permissible to push the one? Yes, but is this the wrong answer? There does not seem to be a morally relevant difference between pushing the one person here and (a) using a trap door which opens to the tracks, or (b) using a "Lazy Susan-type device" (Kamm 1989, 228–229), or (c) throwing the switch in TROLLEY. Hence, there is good reason to claim permissibility of harming the one here, given that it is permissible to throw the switch in TROLLEY (see also below in the main paragraph). Apart from that, one could say in my defense that this is not a very serious counterexample because it already contains some additional complexities and a degree of far-fetchedness such that intuitions are starting to become less and less trustworthy (see, for instance, also the baroque cases in Unger (1996) which should reduce confidence in taking one's "intuitive" judgments very seriously). However, I don't want to use this kind of reply here. A more serious reply in case one doesn't agree with my permissibility verdict would be to invoke the methodology of reflective equilibrium (see above) and argue that the weight should rather be on the side of the principle here than on the side of the particular intuitions. I don't want to pursue this particular methodological point any further here and rather put the main weight on the claim of permissibility.

<sup>20</sup> See also Bruers (2016) on this pair of cases and the general idea of "moral illusions" about relevant moral differences in such cases. For the method of "intermediate cases" used here see Fischer (1992a, b), Fischer and Ravizza (1992a, 1994), Unger (1996, ch. 4), Harris (2000) as well as Boorse and Sorensen (1988) and Boorse (1994) in exchange with Fischer and Ravizza (1994).

<sup>21</sup> The following case would also count as borderline, according to (PEH):

#### SHIPWRECK

Six persons have survived the sinking of their ship. They are all on a lifeboat but close to starving to death. In desperation they consider drawing straws to determine one who will be killed in order to feed the others.

## 6 *Ceteris Paribus*

I said above that (PEH) is defeasible because it is based on certain *ceteris paribus* assumptions. That is, when the *cetera* are not *paria*, the verdict about permissibility can change because (PEH) alone doesn't suffice for a verdict. Consider, for instance this case:

### CRYING BABY

A group of unjustly persecuted people are hiding from the secret police. With them is a baby which is about to start crying. Its parents can either let it cry (which would lead to the discovery and killing of the adults) or suffocate and thus kill the baby (which would rescue the rest of the group and enable them to escape to freedom).

Here, there is neither role-equality nor risk-equality. One reason why one might still judge that killing the baby is permissible has to do with the fact that in a sense the baby poses a threat to the rest of the group (even if only an innocent threat) against which the rest of the group might have some legitimate claim of self-defense.<sup>22</sup>

One could also, on the other hand, imagine a variation on TROLLEY where the five are suicidal (or willing to play Russian Roulette, or better: Russian Trolley) and the one has been lured onto the track by malicious people. In that case we would not be very much willing at all to divert the trolley. If we assume that the organ failures of the five in the transplant-case have been incurred by leading imprudent and reckless lives we would have an independent and additional reason to deny the permissibility of "harvesting" the one (but see Singer 1977, 219–221); similarly, if we invoke a principle according to which people have to put up with life and what it presents to them on their own. On the other hand, if it turns out that the healthy person is in some way directly responsible for the bad health of the five, then it looks more permissible to "harvest" the one.

Here is another case (not involving life or death, for a change):

### SAMARITAN

Three persons are out in the cold. One has a very warm coat that is keeping them warm; the other two are freezing (but not in danger of dying from the

---

Footnote 21 (continued)

Here the individual persons are exchangeable but the profile of gains and losses between the would-be victim and the would-be survivors shows the relevant inequality. For one well-known real life case like that see, e.g., *R v Dudley and Stephens*. If, in addition, the one harmed had special characteristics that uniquely qualified him for being "chosen"—like the fact that their last name is prior in alphabetical order to the other last names—, then there would also be role-inequality and killing the one would be impermissible, according to (PEH).

<sup>22</sup> It also matters whether the baby will survive if discovered or not. If not, then the intuition of permissibility is much stronger because the question rather boils down to whether all in a group or not all in a group should die.—If the situation of the rest of group will improve if not detected (and the baby will survive and get better if detected), then there is (much more) risk equality and the case looks much more like a borderline case.

cold). Should we take the coat from the one and give it to the two who could share it and warm up (with the one starting to freeze)?

Here we have role-equality but no risk-equality (the one cannot get better but the two can get better). Whether forcing the one to be an (unwilling) Samaritan is permissible after all might depend, for instance, on whether or not we think they have a strong antecedent property right in their coat or whether the antecedent distribution of coats was unjust to begin with. Similar but broader questions can be raised concerning the distribution of wealth in general in a given society—for instance about just redistributive tax policies.<sup>23</sup>

Finally and more generally, equality of the kind invoked here might not be the only relevant moral consideration. I want to and (because I have to keep the paper's topic in focus) need to leave the question open here whether, for instance, consequentialist considerations, virtue theoretic, contractualist or deontological considerations of different kinds might play a role when thinking about Trolley-cases, too. If they do, then a further question would be how to balance different types of moral considerations when making judgements about specific cases. This is a problem not just for our views on Trolley-cases but for moral philosophy more generally.

## 7 Conclusion: The Gist of It and Further Questions

So, when the cetera are not paria and additional factors come into the picture, then (PEH) might not suffice for a verdict. It suffices only given the meeting of such ceteris paribus clauses. This dependency on ceteris paribus clauses is typical for moral principles and should neither surprise nor worry us.

Before I come to the very end I would like to mention some questions and topics very much worth pursuing more in the future.<sup>24</sup> As one can already see from the discussion of the different cases and scenarios above, a lot in our judgments about them

<sup>23</sup> Finally, one might want to entertain oneself by considering the following kind of (admittedly) far-fetched case:

### SWITCHING ORGANS

A bystander notices a runaway trolley (whose driver has fainted after brake failure) on the track ahead of which are five people who will not be able to get off the track in time. As the trolley approaches and before it reaches the switch, each of the five suffers a serious shock upon noticing the trolley and consequent organ failure (everyone a different one); if the trolley passes the switch and goes towards the five they all die just before the train hits them. The track has a spur leading off to the right. The bystander can throw a switch thus turning the trolley off the main track and onto the one on the right. Unfortunately there is one person on the right-hand track who will also not be able to get off the track in time. If the trolley hits and kills the one (who also happens to be an organ-donor), then his organs will be taken out and implanted in the five on the other track to save their lives. The bystander can either turn the trolley, bringing about the death of the one; or he can refrain from turning the trolley, bringing about the death of the five.

SWITCHING ORGANS appears to have all the relevant features of both TROLLEY and TRANSPLANT (assuming that it doesn't matter how the trolley kills or how lives are being saved). I can't settle the question here how to judge such cases. For similar cases see also Gert (1993, 175), Russell (1993, 170), Kamm (2001, 182), Richardson (2008, 85).

<sup>24</sup> Thanks to a referee for this suggestion!

depends on the nitty-gritty of the details of the cases, for instance whether someone's situation improves, remains the same, or worsens. More detailed explanations of what counts as better, worse or the same, as well as more detail about what counts as a "role" in "role-equality" could be very useful here. This might help against the risk of succumbing to the temptation of discussing more and more complex, far-fetched or baroque cases in which our verdicts become more and more unstable and uncertain. This would also offer a concrete occasion to revisit questions about reflective equilibrium and other methodological issues concerning our moral judgments (including lessons we can learn from the increasing experimental research).

The principle (PEH) uses two ideas of equality: role-equality and risk-equality. One further important question is whether they always have equal weight or whether their relative weight differs, perhaps even in variation with different kinds of cases (see fn.15). This question is particularly relevant to borderline cases where there is one type of equality but not the other. Getting clearer about relative weights will help much to think about borderline cases.

Another important topic concerns the *cetera* which are often not *paria*. What is morally relevant here? How should we use different kinds of moral considerations (consequentialist, deontological, virtue-ethical, contractualist) in our judgments about Trolley-problems? In the discussion so far, the different options have always been present but they could and should be discussed more explicitly.

The principle proposed here for dealing with the Trolley-problem, (PEH), is attractive because it offers a straightforward response to the Trolley-problem. The basic idea behind it is egalitarian. It explains plausible judgments about cases. And this plausibility is good enough when one is dealing with a tough paradox like the Trolley-problem. Insofar as we subscribe to the value of equality (of the sort addressed by (PEH)), the equality view also offers a good justification for our normative verdicts about the different cases. It goes at least some and—as I like to think—a lot of the way towards a resolution of the Trolley paradox (also see fn.5 for an overview over different specific theoretical options).

**Acknowledgements** For discussions and comments I am very grateful to Reinhard Baule, Gisela Cramer, Paul Rablen, Robert C. Robinson, Bruce Russell, an audience at the 2018 APA Eastern Division Meeting in New York City, an audience at Sungkyunkwan University in Seoul in October 2018, and some referees.

## References

- Bauman, C. W., Peter McGraw, A., Bartels, D. M., & Warren, C. (2014). Revisiting external validity: concerns about trolley problems and other sacrificial dilemmas in moral psychology. *Social and Personality Psychology Compass*, 8(9), 536–554.
- Boorse, C. (1994). Ducking Trolleys. *Journal of Social Philosophy*, 25, 146–152.
- Boorse, C., & Sorensen, R. A. (1988). Ducking harm. *The Journal of Philosophy*, 85, 115–134.
- Bruers, S. (2016). In search of moral illusions. *Journal of Value Inquiry*, 50, 283–303.
- Bruers, S., & Braeckman, J. (2014). A review and systematization of the Trolley problem. *Philosophia*, 42, 251–269.
- Cathcart, T. (2013). *The Trolley problem; Or, would you throw the fat guy off the bridge? A philosophical conundrum*. New York: Workman.

- Clark, M. (1995). Sacrificing one to save many. *Journal of Applied Philosophy*, 12, 189–200.
- Costa, M. J. (1986). The Trolley problem revisited. *The Southern Journal of Philosophy*, 24(4), 437–449.
- Costa, M. J. (1987). Another trip on the Trolley. *The Southern Journal of Philosophy*, 25(4), 461–466.
- Cushman, F., Young, L., & Hauser, M. (2006). The role of conscious reasoning and intuition in moral judgment: Testing three principles of harm. *Psychological Science*, 17(12), 1082–1089.
- Di Nucci, E. (2013). Self-sacrifice and the Trolley problem. *Philosophical Psychology*, 26(5), 662–672.
- Edmonds, D. (2014). *Would you kill the fat man? The Trolley problem and what your answer tells us about right and wrong*. Princeton: Princeton University Press.
- Fischer, J. M. (1992a). The Trolley and the sorites. *Yale Journal of Law and the Humanities*, 4, 105–126.
- Fischer, J. M. (1992b). Thoughts on the Trolley problem. In J. M. Fischer & M. Ravizza (Eds.), *Ethics: Problems and principles* (pp. 310–317). Fort Worth: Harcourt Brace Jovanovich.
- Fischer, J. M., & Ravizza, M. (1992a). Thomson and the Trolley. *Journal of Social Philosophy*, 23, 64–87.
- Fischer, J. M., & Ravizza, M. (1992b). Introduction. In J. M. Fischer & M. Ravizza (Eds.), *Ethics: Problems and principles* (pp. 1–57). Harcourt Brace Jovanovich: Fort Worth.
- Fischer, J. M., & Ravizza, M. (1994). Ducking harm and sacrificing others. *Journal of Social Philosophy*, 25(3), 135–145.
- Foot, P. (Ed.). (1978). The problem of abortion and the doctrine of the double effect. In *Virtues and vices and other essays in moral philosophy* (pp. 19–32). Oxford: Clarendon.
- Fried, B. H. (2012). What does matter? The case for killing the Trolley problem (or letting it die). *The Philosophical Quarterly*, 62, 505–529.
- Frowe, H. (2018). Lesser-evil justifications for harming: Why we're required to turn the Trolley. *The Philosophical Quarterly*, 68, 460–480.
- Gert, B. (1993). Transplants and Trolleys. *Philosophy and Phenomenological Research*, 53, 173–179.
- Greene, J. D., Brian Sommerville, R., Nystrom, L. E., Darley, J. M., & Cohen, J. D. (2001). An fMRI investigation of emotional engagement in moral judgment. *Science*, 293, 2105–2108.
- Griffin, J. (1996). *Value judgment. Improving our Ethical Beliefs*. Oxford: Clarendon.
- Hallborg, R., Jr. (1997). The lesser-evil defense and the Trolley problem. *Legal Theory*, 3, 291–316.
- Hanna, R. (1992). Morality De Re: reflections on the Trolley problem. In J. M. Fischer & M. Ravizza (Eds.), *Ethics: Problems and principles* (pp. 320–336). Harcourt Brace Jovanovich: Fort Worth.
- Hare, R. M. (1981). *Moral thinking. Its levels, method, and point*. Oxford: Clarendon.
- Harris, J. (2000). The moral difference between throwing a Trolley at a person and throwing a person at a Trolley. *Proceedings of the Aristotelian Society, Supplementary Volumes*, 74, 41–57.
- Haslett, D. W. (2011). Boulders and Trolleys. *Utilitas*, 23, 268–287.
- Hauser, M., Cushman, F., Liane Young, R., Jin, K.-X., & Mikhail, J. (2007). A dissociation between moral judgments and justifications. *Mind and Language*, 22, 1–21.
- Ishiguro, K. (2005). *Never let me go*. London: Faber and Faber.
- Kamm, F. M. (1989). Harming some to save others. *Philosophical Studies*, 57, 227–260.
- Kamm, F. M. (2000). The doctrine of triple effect and why a rational agent need not intend the means to his end. *Proceedings of the Aristotelian Society*, 74, 21–39.
- Kamm, F. M. (2001). *Morality, mortality: Volume II: Rights, duties, and status*. Oxford: Oxford University Press.
- Kamm, F. M. (2016a). Lecture I: Who turned the Trolley? In F. M. Kamm & E. Rakowski (Eds.), *The Trolley problem mysteries* (pp. 11–56). Oxford: Oxford University Press.
- Kamm, F. M. (2016b). Lecture II: How was the Trolley turned? In F. M. Kamm & E. Rakowski (Eds.), *The Trolley problem mysteries* (pp. 57–102). Oxford: Oxford University Press.
- Lanteri, A., Chelini, C., & Rizzello, S. (2008). An experimental investigation of emotions and reasoning in the Trolley problem. *Journal of Business Ethics*, 83(4), 789–804.
- Laycock, D. (1985/86). The ultimate unity of rights and utilities. *Texas Law Review* 64, 407–413.
- Liao, S. Matthew, Wiegmann, A., Alexander, J., & Vong, G. (2012). Putting the Trolley in order: Experimental philosophy and the loop case. *Philosophical Psychology*, 25(5), 661–671.
- Locke, D. (1982). The choice between lives. *Philosophy*, 57, 453–475.
- Mikhail, J. (2011). *Elements of moral cognition*. Cambridge: Cambridge University Press.
- Montmarquet, J. A. (1982). On doing good: The right and the wrong way. *The Journal of Philosophy*, 79, 439–455.
- Navarrete, C. David, McDonald, M. M., Mott, M. L., & Asher, B. (2012). Virtual morality: Emotion and action in a simulated three-dimensional “Trolley Problem”. *Emotion*, 12, 364–370.

- Naylor, M. B. (1988). The Moral of the Trolley Problem. *Philosophy and Phenomenological Research*, 48, 711–722.
- Norcross, A. (2008). Off her trolley? Frances Kamm and the metaphysics of morality. *Utilitas*, 20(1), 65–80.
- Petrinovich, L., & O'Neill, P. (1996). Influence of wording and framing effects on moral intuitions. *Ethology and Sociobiology*, 17, 145–171.
- Rai, T. S., & Holyoak, K. J. (2010). Moral principles or consumer preferences? Alternative Framings of the Trolley Problem. *Cognitive Science*, 34, 311–321.
- Richardson, H. (2008). Discerning Subordination and Inviolability: A Comment on Kamm's. *Intricate Ethics Utilitas*, 20, 81–91.
- Russell, B. (1977). The relative strictness of positive and negative duties. *American Philosophical Quarterly*, 14, 87–97.
- Russell, B. (1993). Exploring the realm of rights. *Philosophy and Phenomenological Research*, 53, 169–172.
- Schmidtz, D. (2006). *The elements of justice*. Cambridge: Cambridge University Press.
- Shaw, J. (2006). Intentions and Trolleys. *The Philosophical Quarterly*, 56, 63–83.
- Sheng, C. L. (1995). A suggested solution to the Trolley problem. *Journal of Social Philosophy*, 26(1), 203–217.
- Singer, P. (1977). Utility and the survival lottery. *Philosophy*, 52, 218–222.
- Stark, A. (2017). The Trolley's last stop before consequentialism: Exploring the terrain. *Ethical Theory and Moral Practice*, 20, 1021–1035.
- Taurek, J. (1977). Should the numbers count? *Philosophy and Public Affairs*, 6, 293–316.
- Terbeck, S., Kahane, G., McTavish, S., Savulescu, J., Levy, N., Hewstone, M., et al. (2013). Beta adrenergic blockade reduces utilitarian judgement. *Biological Psychology*, 92, 323–328.
- Thomson, J. J. (1976). Killing, letting die, and the Trolley problem. *The Monist*, 59, 204–217.
- Thomson, J. J. (1985). The Trolley problem. *The Yale Law Journal*, 94, 1395–1415.
- Thomson, J. J. (1990). *The realm of rights*. Cambridge, MA: Harvard University Press.
- Thomson, J. J. (2008). Turning the Trolley. *Philosophy and Public Affairs*, 36, 359–374.
- Thomson, J. J. (2016). Kamm on the Trolley problems. In F. M. Kamm & E. Rakowski (Eds.), *The Trolley problem mysteries* (pp. 113–134). Oxford: Oxford University Press.
- Unger, P. (1996). *Living high and letting die*. Oxford: Oxford University Press.
- Waldmann, M. R., & Dieterich, J. H. (2007). Throwing a bomb on a person versus throwing a person on a bomb: Intervention myopia in moral intuitions. *Psychological Science*, 18(3), 247–253.
- Wiech, K., Kahane, G., Shackel, N., Farias, M., Savulescu, J., & Tracey, I. (2013). Cold or calculating? Reduced activity in the subgenal cingulate cortex reflects decreased emotional aversion to harming in counterintuitive utilitarian judgment. *Cognition*, 126, 364–372.
- Williams, B. (1973). A critique of utilitarianism. In J. J. C. Smart & B. Williams (Eds.), *Utilitarianism for and against* (pp. 77–150). Cambridge: Cambridge University Press.
- Wood, A. (2011). Humanity as end in itself. In D. Parfit & S. Scheffler (Eds.), *On what matters* (Vol. 2, pp. 58–82). Oxford: Oxford University Press.

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.