Abstract Judith Jarvis Thomson has recently proposed a new argument for the thesis that killing the one in the Trolley Problem is not permissible. Her argument relies on the introduction of a new scenario in which the bystander may also sacrifice herself to save the five. Thomson argues that those not willing to sacrifice themselves if they could may not kill the one to save the five. Bryce Huebner and Marc Hauser have recently put Thomson’s argument to the empirical test by asking people what they should do in the new trilemma case in which they may also sacrifice themselves. They found that the majority judge that they should either kill the one or sacrifice themselves; Huebner and Hauser argue that those numbers speak against Thomson’s argument. But Thomson’s argument was about the dialectical effect of the new trilemma on the traditional dilemma, rather than about the trilemma itself. Here I present the results of a study in which I asked subjects first what they should do in the trilemma and then what they should do in the traditional Trolley Problem. I found that, if asked first about the trilemma, subjects then have the intuition that killing the one in the traditional Bystander at the Switch is not permissible – exactly what Thomson’s argument had predicted.

In one of the infamous thought-experiments of analytic philosophy, a runaway trolley is about to kill five workmen who cannot move off the tracks quickly enough; their only chance is for a bystander to flip a switch to divert the trolley onto a side-track, where one workman would be killed. In a parallel scenario, the bystander’s only chance to save the five is to push a fat man off a bridge onto the tracks: that will stop the trolley but the fat man will die. Why is it permissible for the bystander to divert the trolley onto
the one workman by pressing the switch while it is not permissible for the bystander to
stop the trolley by pushing the fat man off the bridge?

This is the so-called Trolley Problem, resulting from Judith Jarvis Thomson’s
(1976 & 1985) adaptation of an example originally devised by Philippa Foot (1967). If
it is permissible to intervene in the so-called Bystander at the Switch scenario while it is
not permissible to intervene in the so-called Fat Man scenario, then the Trolley Problem
arises and we must explain the moral difference between these two cases.

According to the Doctrine of Double Effect (DDE), “sometimes it is permissible
to bring about as a merely foreseen side effect a harmful event that it would be
impermissible to bring about intentionally”, as McIntyre puts it in her introductory
essay on DDE for the Stanford Encyclopedia of Philosophy.\(^1\) This is in line with classic
formulations of the Doctrine such as Mangan’s: “A person may licitly perform an action
that he foresees will produce a good and a bad effect provided that four conditions are
verified at one and the same time: 1) that the action in itself from its very object be good
or at least indifferent; 2) that the good effect and not the evil effect be intended; 3) that
the good effect be not produced by means of the evil effect; 4) that there be a
proportionately grave reason for permitting the evil effect” (1949: 43).\(^2\)

DDE offers a possible answer to the Trolley Problem, because it can be
deployed to argue that the difference in moral permissibility results from the one being
killed as a means to saving the five in Fat Man; while in Bystander at the Switch the
killing of the one is a mere side-effect of saving the five. In this respect, as long as the
Trolley Problem remains “unsolved” it offers dialectical support to DDE. So the
connection between DDE and the Trolley Problem is dialectically very simple: the
Trolley Problem counts as an argument in favour of DDE in so far as it remains an
unresolved problem and in so far as the Doctrine offers a possible solution to this
unresolved problem. On top of this, the strong intuitions behind the Trolley Problem
suggest that DDE is intuitively appealing and possibly even, looking past normative
ethics, that it is deeply embedded in our moral psychology.

Intuitions on the Trolley Problem are, indeed, very clear. If the results of Marc
Hauser’s Moral Sense Test are to be believed, then according to public opinion it is

\(^1\) [http://plato.stanford.edu/entries/double-effect/](http://plato.stanford.edu/entries/double-effect/)

\(^2\) Mangan follows Gury, which is the standard modern Latin source: “It is lawful to actuate a morally
good or indifferent cause from which will follow two effects, one good and the other evil, if there is a
proportionately serious reason, and the ultimate end of the agent is good, and the evil effect is not the
means to the good effect” (Mangan’s translation; 1949: 60). Aquinas is often credited with the first
explicit version of DDE: “Nothing hinders one act from having two effects, only one of which is
intended, while the other is beside the intention” (Summa II-II, 64, 7).
indeed permissible to intervene in Bystander at the Switch: around 90% of respondents to the Moral Sense Test thought as much (Hauser 2006: 139). On the other hand, respondents say that it is not permissible to intervene in Fat Man: only around 10% of respondents thought it permissible to intervene.³

The folk, then, are due an answer: why is it permissible to intervene in Bystander at the Switch while it is not permissible to intervene in Fat Man? The literature on DDE and the Trolley Problem is very large, because the question drives straight at the heart of the disagreement between deontological and consequentialist approaches to normative ethics. There are three possible approaches here: (1) one can try to answer the question; or (2) deny the question on the grounds that it is permissible to kill the fat man; or (3) deny the question on the grounds that it is not permissible to kill the one in Bystander at the Switch. Here I will ignore both 1 and 2: on 2, biting the consequentialist bullet and denying that it is not permissible to kill the fat man is, whatever the merits of such a strategy, not a very original way to go: these kinds of counterexamples have always been brought against consequentialism. On 1, answers to the question have flooded the literature in the last thirty years, but here I won’t analyse or even just survey those⁴. I will rather focus on 3, the claim that it is not permissible to intervene in Bystander at the Switch, and that therefore we don’t need to answer the question that has come to be known as the Trolley Problem simply because there is no such problem.

Thomson herself (2008) has recently argued that the Trolley Problem is not actually a problem because it is not permissible to divert the trolley to kill the one workman in Bystander at the Switch. Thomson introduces a new scenario in which the bystander has also the possibility of sacrificing herself to save the five workmen. Here is Thomson’s new Bystander’s Three Options:

(i) do nothing, letting five die, or
(ii) throw the switch to the right, killing one, or
(iii) throw the switch to the left, killing himself (2008: 364).

³ A BBC online survey had broadly similar, if not so dramatic, results: 76,85% said it was permissible to intervene in Bystander at the Switch, while 73,12% said it was not permissible to intervene in Fat Man (http://news.bbc.co.uk/2/hi/uk_news/magazine/4954856.stm).
Thomson thinks that in this scenario choosing (ii) would not be right: “I hope you will agree that choosing (ii) would be unacceptable on the bystander’s part. If he can throw the switch to the left and turn the trolley onto himself, how dare he throw the switch to the right and turn the trolley onto the one workman?” (2008: 364). More importantly, Thomson thinks that the new scenario also illustrates why the bystander may not kill the one in the traditional Bystander at the Switch: “Since he wouldn’t himself pay the cost of his good deed if he could pay it, there is no way in which he can decently regard himself as entitled to make someone else pay it” (2008: 366). Thomson claims, then, that if the bystander would not be willing to sacrifice herself to save the five, then she may not sacrifice someone else.

Thomson’s is, then, an argument about the traditional Bystander at the Switch: her new scenario is just the hypothetical question that, according to Thomson, the bystander should ask itself in deciding whether she may intervene. And Thomson claims that if the bystander cannot answer to the hypothetical question that she would indeed sacrifice herself, then she may not intervene.

Bryce Huebner and Marc Hauser have put Thomson’s new trilemma to the experimental test. They have asked participants to the online Moral Sense Test the following question:

Jesse is standing near the railroad tracks and notices an empty boxcar coming down the tracks, moving fast enough to kill anyone that it hits. If Jesse does nothing, the boxcar will continue along the main track, killing five people who are walking down the main track. There is a switch nearby that Jesse can use to divert the boxcar onto either of two side tracks that split off from the main track in opposite directions. There is one person walking along the right-side track. So, if Jesse flips the switch to the right, the boxcar will hit and kill this person. Jesse’s foot is stuck in the track on the left-side track. So if Jesse flips the switch to the left, he will be hit and killed by the trolley himself. What should Jesse do? (2011: 95)

The answers were as follows:

- 43% judged that Jesse should flip the switch to the right;
- 38% judged that Jesse should sacrifice herself;
- 18.7% judged that Jesse should allow the trolley to proceed along the main track.

Page numbers refer to the pre-print available here: http://www9.georgetown.edu/faculty/lbh24/MJAASS.PDF
In a second version, the same question was put to respondents with the only variant that ‘Jesse’ was replaced with the pronoun ‘you’. The outcome was very similar: 48% judged that they should flip the switch to the right; 33,7% judged that they should sacrifice themselves; and 18,2% judged that they should do nothing.

Huebner and Hauser think that their numbers speak against Thomson’s hypothesis: “The data reported in the previous section minimally suggests that Thomson’s intuition is radically at odds with the commonsense intuition of what a person should do when she is faced with a difficult moral trilemma such as the three-track bystander case. If Thomson were right, we would expect it to be far more transparent that it is immoral to turn the trolley, and this increased transparency should be reflected, at least to some extent, in the folk-moral judgments that are offered in response to this case. Even if not everyone was sensitive to the hypocrisy of turning the trolley, we would expect a striking increase in the proportion of people who judge that one should just let the trolley go in such a case. In stark contrast to this hypothesis, we found no such pattern in folk-moral judgments about the three-track bystander case. Across both conditions in which altruistic self-sacrifice was at issue, only a small minority of participants (approximately 18%) judged that they should let the trolley go, suggesting that even when people are explicitly presented with the option of altruistic self-sacrifice, the inclination to save the five people on the main track continues to dominate their intuitive judgments” (2011: 13-14).

I think that this is a reasonable interpretation of the numbers: here I don’t wish to deny the significance of those numbers nor do I want to offer alternative interpretations of those numbers. I would rather like to point out that Huebner and Hauser have not, even by their own standards, asked the right question; or, at least, that there was, even by their standard, a better question to ask. There was a much more significant question to ask, and this emerges quite clearly from their own discussion: “Thomson’s philosophical intuition is that considering the option of altruistic self-sacrifice makes the negative duty not to harm another person more salient. If this is the case, then perhaps the introduction of such an option will also have a significant effect on the proportion of people who will judge that it is permissible to flip the switch in a trolley case” (2011: 8).

Huebner and Hauser correctly identify the dialectic of Thomson’s argument: her new scenario is only significant in so far as it illustrates (or makes more salient, to use their expression) something about the traditional trolley problem. But, then, why did Huebner and Hauser not ask participants who took the trilemma about the traditional trolley problem? If what Huebner and Hauser wanted to test Thomson’s argument on
was whether, as she claims, the trilemma has an effect on the traditional trolley problem, why not ask indeed about the traditional trolley problem?

“Thomson assumes, based on her own intuitions, that the introduction of altruistic self-sacrifice should lead others to see the hypocrisy inherent in sacrificing the person on the side track and to judge that the bystander should let the trolley continue along the main track” (2011: 7). This is exactly what Thomson claims: and, importantly, she does so not just in the case of the new trilemma, but also in the case of the traditional trolley problem. But then the only way of testing Thomson’s claim is to ask participants about the traditional trolley problem.

This is exactly what I did, finding that the trilemma significantly influences answers to the traditional dilemma. In December 2011 152 subjects took an online survey that I designed by combining Thomson’s trilemma (in the very wording of Huebner and Hauser’s version) and the traditional Bystander at the Switch. Participants were randomly presented with either

- Thomson’s trilemma and then the traditional dilemma; or with

- the traditional dilemma and then Thomson’s trilemma; or with

- a trilemma not involving self-sacrifice (also modeled on one presented by Huebner and Hauser) and then the traditional dilemma.

Participants did not have access to the second question while answering the first, and they could not change their answer to the first question after having looked at or answered the second question. For the trilemma, I used Huebner and Hauser’s Condition 2 question, where the pronoun ‘you’ replaces ‘Jesse’: I did so on the grounds of the principle of charity, as with ‘you’ the third option of doing nothing (that one that supposedly would have supported Thomson’s argument) had even worse numbers than in Condition 1 with ‘Jesse’: 18,2% instead of 18,7%.

The trilemma question read as follows:

1) You are standing near the railroad tracks and notice an empty boxcar coming down the tracks, moving fast enough to kill anyone that it hits. If you do nothing, the boxcar will continue along the main track, killing five people who are walking down the main track. There is a switch nearby that you can use to divert the boxcar onto either of two side tracks that split off from the main track in opposite directions. There is one person walking along the right-side track. So, if you flip the switch to the right, the boxcar will
hit and kill this person. Your foot is stuck in the track on the left-side track. So if you flip the switch to the left, you will be hit and killed by the trolley yourself. What should you do?

The three possible answers, whose order of presentation to the respondent was randomized, were:

- I do nothing
- I flip the switch to the right
- I flip the switch to the left

The traditional dilemma question was designed to replicate the Bystander at the Switch scenario while at the same time maintaining the wording of the trilemma question. It read as follows:

2) You are standing near the railroad tracks and notice an empty boxcar coming down the tracks, moving fast enough to kill anyone that it hits. If you do nothing, the boxcar will continue along the main track, killing five people who are walking down the main track. There is a switch nearby that you can use to divert the boxcar onto a side-track where one person is walking. So, if you flip the switch, the boxcar will hit and kill this person. What should you do?

The possible answers, whose order of presentation to the respondent was randomized, were:

- I do nothing
- I flip the switch

The third condition, designed in order to control for the specific effect of self-sacrifice and not just for the effect of any trilemma preceding the traditional dilemma, read as follows:

3) You are standing near the railroad tracks and notice an empty boxcar coming down the tracks, moving fast enough to kill anyone that it hits. If you do nothing, the boxcar will continue along the main track, killing five people who are walking down the main track. There is a switch nearby that you can use to divert the boxcar onto either of two side tracks that split off from the main track in opposite directions. There is one person
walking along the right-side track. So, if you flip the switch to the right, the boxcar will hit and kill this person. There is another person walking along the left-side track. So if you flip the switch to the left, the boxcar will hit and kill this person. What should you do?

The three possible answers, whose order of presentation to the respondent was randomized, were:

- I do nothing
- I flip the switch to the right
- I flip the switch to the left

The three different questions were combined to form three conditions:

a. 1, then 2  
b. 2, then 1  
c. 3, then 2

Conditions (a) and (b) tested the effect of the trilemma on the traditional dilemma; condition (c) tested the specific effect of self-sacrifice as opposed to the general effect of a trilemma before answering the traditional dilemma. 152 subjects participated: 75 answered condition (a); 45 answered condition (b); 32 answered condition (c).

The answers to the dilemma question in the three conditions were as follows:

Condition (a): flip the switch 38,66%; not flip the switch 61,34%;

Condition (b): flip the switch 66,67%; not flip the switch 33,33%;

Condition (c): flip the switch 56,25%; not flip the switch 43,75%.

---

6 The results are statistically significant: $\chi^2(2, N: 152) = 9,31$, $p<0.01$
7 Answers to the trilemma for condition (a): do nothing 37, flip to the right 26, and flip to the left 11.
8 Answers to the trilemma for condition (b): do nothing 14, flip to the right 15, and flip to the left 13.
9 Answers to the trilemma for condition (c): do nothing 11, flip to the right 9, and flip to the left 11.
The difference between the results of Bystander at the Switch when answered in isolation and Bystander at the Switch when answered after having answered Thomson’s trilemma is impressive.\textsuperscript{10} In the former case, a significant majority answer that intervening is permissible (confirming previous studies). In the latter case, the intuition that intervening is permissible disappears, with just over a third of respondents sharing it. Importantly, my study suggests that they reach such conclusion only after having been subjected to the self-sacrifice trilemma, just as Thomson argued: when answering the traditional dilemma only after having answered a trilemma that did not involve self-sacrifice, the majority still has the intuition that flipping the switch is permissible.

Recall the words of Huebner and Hauser: “Thomson’s philosophical intuition is that considering the option of altruistic self-sacrifice makes the negative duty not to harm another person more salient. If this is the case, then perhaps the introduction of such an option will also have a significant effect on the proportion of people who will judge that it is permissible to flip the switch in a trolley case” (2011: 8). In my study the introduction of such an option did indeed have a significant effect on the proportion of

\textsuperscript{10} The results are statistically significant: $\chi^2 (1, N: 120)= 8.82$, $p<0.01$
people who judge that it is permissible to flip the switch in a trolley case: those people drastically dropped from 66.67% down to 38.66%. And this is a direct result of considering the option of altruistic self-sacrifice, exactly as Thomson had predicted in her theoretical argument.

It may be objected that more needs to be done to prove that these results support Thomson’s argument; my results only show, it may be argued, that the relevant moral intuitions are very unstable. This is after all the conclusion Swain et al. (2008) drew following their own experimental philosophy studies where the order in which the cases were presented affected results (see also, for other examples of this sort of order effect, Petrinovich & O’Neill 1996; Sinnott-Armstrong 2008; Wiegmann et al. 2010). I think that the dialectic of Thomson’s argument is such that our results do support it, because her argument is about the effect of the new case on the old one (and Huebner and Hauser seem to think that too). Also, condition (c) controls exactly against just any order effect: it isn’t simply the introduction of a question before the traditional dilemma; and it isn’t even the introduction of a trilemma before the traditional dilemma; it is the introduction of the option of self-sacrifice.\footnote{This claim needs to be qualified: while there is no statistically significant difference between (b) and (c) ($\chi^2 (1, N: 77)= 0.87, p<0.01$), which suggests that a trilemma which does not involve self-sacrifice has no significant effect on the answers to the traditional dilemma, the chi-square between (a) and (c) also fails to reach statistical significance ($\chi^2 (1, N: 107)= 2.82, p<0.01$), weakening the claim that it is exactly the self-sacrifice element which induces subjects to then not kill the one in the dilemma.}

Let me say something in the way of explaining the results of my experiment: why are the answers to the traditional Trolley Problem affected by engaging with the self-sacrifice trilemma? Here there are at least two alternative ways of thinking: (1) one may take it that having looked at the trilemma shifts the attention of the respondent, when thinking afterwards about the traditional dilemma, to the similarity between the one workman and oneself. The idea would be that the trilemma helps to let the respondent identify with the one workman: they are both, after all, on a side-track towards which a threat may be redirected. This idea of a shift of perspective enabled or facilitated by the trilemma may indeed explain the results, but would it also support the normative conclusion that killing the one in the traditional Trolley Problem is indeed impermissible? Alternatively, (2) it may be argued that engaging with the trilemma does not just shift the agent’s perspective, but actually allows the agent to see the morally relevant features which make intervening in the traditional Trolley Problem impermissible. It is only by comparing the one workman with ourselves that we realize the true nature of the workman’s situation: namely that we are asking him to die for the sake of strangers.
Let me emphasize that the significance of my experiment does not depend on being able to defend the normative value of the effect that I have identified. Even if one wants to dispute the effect of the trilemma on the dilemma, our results would still show, in line with Swain et al. (2008), that the Bystander intuition is unstable and that it depends on the order of presentation so heavily that it disappears when Bystander is first introduced after Thomson’s new self-sacrifice case. This latter outcome is anyhow more important than supporting Thomson’s argument, because it is significantly more general in two different respects: (1) we must distinguish between the claim that killing the one in Bystander at the Switch is not permissible and arguments, such as Thomson’s, for such claim that killing the one is not permissible. My results are relevant to the claim that killing the one is not permissible whether or not they support Thomson’s argument. This is because my results defuse the supposed overwhelming intuition that it is permissible to kill the one. (2) There is an aspect of Thomson’s argument that we have not yet emphasized: if her argument is that if you are not willing to sacrifice yourself you may not sacrifice someone else, then this argument – in such a form – does not have anything to say about those who are willing to sacrifice themselves. Indeed, one would suppose that, given Thomson’s reasoning, those who are willing to sacrifice themselves would be then allowed to sacrifice the one. This brings us back to the one workman and dying for the sake of strangers.

Thomson tries to argue that even those who would be willing to sacrifice themselves may not kill the one. Thomson argues on two grounds: (a) she says that “altruism that rises to this level is not morally attractive. Quite to the contrary. A willingness to give up one’s life simply on learning that five others will live if and only if one dies is a sign of a serious moral defect in a person. “They’re my children,” “They’re my friends,” “They stand for things that matter to me,” “They’re young, whereas I haven’t much longer to live,” “I’ve committed myself to doing what I can for them”: these and their ilk would make sacrificing one’s life to save five morally intelligible. Consider, by contrast, the man who learns that five strangers will live if and

---

12 Here there are other theoretical options that Thomson does not discuss: it could be argued, for example, that it is not enough for the bystander to ask herself whether she would be willing to sacrifice her own life and then, in case she gives herself an affirmative answer, then she may kill the one. Maybe the bystander must know that she would sacrifice herself, or at least she must be justified in believing that she would kill herself – and introspection may not be the right sort of justification here. In general, it can be argued that the epistemic burden on the bystander, in order to be allowed to kill the one in the traditional dilemma, is much heavier than just a Yes or No answer to the relevant hypothetical question.

13 Here my discussion will be very brief, but I analyze at length this and related issues elsewhere: Di Nucci (submitted a), Di Nucci (submitted b), Di Nucci (submitted c), Di Nucci (submitted d), and Di Nucci (submitted e).
only if they get the organs they need, and that his are the only ones that are available in time, and who therefore straightway volunteers. No reputable surgeon would perform the operation, and no hospital would allow it to be performed under its auspices. I would certainly not feel proud of my children if I learned that they value their own lives as little as that man values his” (2008: 366-67); (b) Thomson says that “It remains the case that the altruistic bystander is not entitled to assume that the one workman is equally altruistic, and would therefore consent” (2008: 367).

Thomson’s second point is quickly shown to be problematic: one cannot appeal to consent in arguments about the permissibility of intervening in the Trolley Problem quite simply because what is at stake is that intervention is permissible even in the absence of consent. So this approach will not do, even though it may be interesting to enquire whether those that do think that intervening is permissible may take the one to be obliged to consent (so-called ‘normative consent’) or to implicitly give her consent given the circumstances (so-called ‘tacit consent’). What about Thomson’s other approach, the idea that radical altruism is not attractive? This approach may be more promising, but it has to be supplemented with the claim that the kind of radical altruism involved in dying for the sake of five strangers is not morally valuable. If it were not morally valuable, then that would mean that someone who would be willing to sacrifice herself for the sake of five strangers would have to appeal to liberty – and since liberty as a justification would not apply to the case of killing the one, then we would have an argument for the claim that killing the one in Bystander at the Switch is not permissible on the grounds that radical altruism is not morally valuable.

This is, admittedly, a radical moral stance to uphold; but, importantly, I do not need to defend it here: the point of this paper was to empirically test Thomson’s hypothesis that the trilemma would have an effect on the traditional Bystander at the Switch dilemma. The results are overwhelming: when answering the traditional Trolley Problem only after having answered Thomson’s self-sacrifice trilemma, a large majority responds that it is not permissible to kill the one.

14 On this and related issues, see FitzPatrick 2009, Shaver 2011, and Di Nucci (submitted a).
15 Many thanks to audiences in Eindhoven and Essen, to Karoline Sachs, and to two anonymous referees for this journal.
References

Di Nucci, E. (submitted a), The Doctrine of Double Effect and the Trolley Problem.
Di Nucci, E. (submitted b), Double Effect and Terror Bombing.
Di Nucci, E. (submitted d), Fat Terror Bombers and Strategic Bystanders.
Di Nucci, E. (submitted e), Intention, Causal Beliefs, and Terror Bombing.


