

The Trolley Problem, Sacrifice, and Self-Defense*

Thomas Nadelhoffer

*Draft Copy: Please do not cite without permission

Introduction:

Philosophers have spilled a lot of ink trying to explain and justify their intuitions with respect to a family of famous thought experiments that are often collectively referred to as “the trolley problem.” For present purposes, I am going to limit my attention to two of these thought experiments in particular—namely, Judith Jarvis Thomson’s Bystander at the Switch and Bystander on the Footbridge cases. My goal in this paper is two-fold: First, I will argue that when we view these two cases from the perspective of the persons whose lives are at stake rather than from the point of view of the bystander, we find that the cases are more symmetrical than Thomson and others have assumed. Second, I will try to show that if a bystander decides to kill one person in order to save five—either by diverting a trolley or by pushing a large man off a footbridge—the person whose life would thereby be placed in danger has a legitimate right to self-defense in either case. Looking at the trolley problem against the backdrop of self-defense not only brings to the surface important issues that need to be taken into consideration before we reach any definitive conclusions concerning how best to solve the problem, it also highlights a tension that exists more generally between the permissibility conditions for self-defense and the permissibility conditions for sacrificing people against their wishes in the name of the greater good.

1. Setting the Stage

Perhaps the most natural way for us to begin our investigation is to simply look at the two thought experiments that will serve as the focus of this essay:

Bystander at the Switch (BAS)

Suppose that you have been strolling by a trolley track when you suddenly realize that there is a runaway trolley coming in your direction. It is clear if the trolley continues on its present path, it will kill five track workmen who are repairing the track up ahead. However, you notice that there is a switch immediately in front of you that you could hit that would turn the trolley onto a side track off to the right. Unfortunately, there is one track workmen on the side track who would certainly be killed if you intervene by hitting the switch and diverting the train. Is it morally permissible for you to intervene by hitting the switch—thereby intentionally killing one in order to save five others?

Bystander on the Footbridge (BOF)

Suppose that you are standing on a footbridge that transcends over a trolley track when you suddenly realize that there is a runaway trolley coming in your direction. It is clear that if the trolley continues on its present path, it will kill five track workmen who are repairing the track up ahead. Luckily, you are an expert on trolleys, so you know that the best way to stop a runaway trolley is to drop something really heavy in its path. And it just so happens that there is a very large man standing next to you on the footbridge who is leaning over the railing watching the trolley as it rapidly approaches. You realize that if you just give him a shove, you can use his body to stop the trolley from killing the five workmen. Is it morally permissible for you to intervene by intentionally pushing the large man to his death in order to save five?

According to Thomson, while it is *impermissible* (and inexcusable) for the bystander to push the large man off the footbridge in BOF (1991: 1409), it is *permissible* for the bystander to hit the switch in order to divert the train onto the side track in BAS (1991: 1395).¹ Given that the same numbers game plays out in both cases—namely, in order to save five people the bystander must knowingly bring about the death of one person—what is it about BAS that makes intervention permissible that does not at the same time make it permissible for the bystander to intervene in BOF?

Before we examine how Thomson attempts to answer this question, I first want to highlight some features that BAS and BOF have in common:

1. Both cases involve an innocent bystander who is not responsible for the events that are unfolding but who nevertheless has the ability to intervene in order to save five workmen.
2. Each bystander has a choice between passively allowing five to die and actively bringing about the death of one.

¹ Thomson does not claim that the bystander in BAS is morally obligated to hit the switch, she merely claims that it is morally permissible for her to do so. To say that x-ing is morally permissible in a given context is to say that an agent would be justified in x-ing in that context. To say that x-ing is morally obligatory in a given context, on the other hand, is to say that non-x-ing would be morally impermissible in that context.

3. There is no special bond that ties the bystanders to the other characters in the two respective thought experiments—e.g., family, friendship, romance, etc.
4. None of the workmen waive their general right to life just because they voluntarily perform a job that involves the risk of loss of life or limb.
5. The bystanders don't have any ill will towards the individuals whose lives they have decided to sacrifice nor do they owe any debts to them.
6. There is no salient moral difference between each of the five workmen, on the one hand, and either the lone workman or the large man, on the other hand.
7. Both of the bystanders would avoid killing anyone at all if they could—but as things stand, the only way for the bystanders to save five is to adopt a plan that involves sacrificing one.
8. By intentionally hitting the switch or pushing the large man, the bystander knowingly adopts a plan that will harm and wrong one person in order to save five others.
9. By knowingly adopting a plan that will harm and wrong one in order to save five, the bystander violates the general right to life of the person whose life she decides to sacrifice.
10. In both cases, the person whose life the bystander decides to sacrifice does not waive her general right to life in the name of the greater good.

These are similarities between BAS and BOF that Thomson either explicitly acknowledges or presumably would acknowledge if pressed to do so. Consequently, if she wants to justify her intuition that it is permissible for the bystander in BAS to hit the switch but that it is neither permissible nor excusable for the bystander in BOF to push the large man off the bridge, Thomson will need to identify a difference between the two scenarios that is robust enough to do the argumentative heavy lifting.²

2. The Trolley Problem Revisited

According to Thomson, in order to solve the Trolley Problem we must focus primarily on the notion of rights as it relates to the means-end relationship between the bystanders and those whom they would harm in their efforts to save the five. More specifically, she suggests that despite the fact that in both BAS and BOF the bystander *wrongs* the person

² The difference between a justification and an excuse is that actions that are merely excusable are wrong even if the agents who perform them are without fault. For instance, when a murderer suffers from a sufficiently debilitating mental disorder, he should not be held responsible for his actions even though they were wrong, morally speaking. In cases of justified self-defense, on the other hand, the killing is permissible and not merely excusable. It is not that I am wrong but faultless for killing an aggressor in self-defense; rather, in some important respects, I have done the *right* thing (or at least something that I had a *right* to do).

whose life she sacrifices, only BOF involves a *direct* violation of someone's rights. On this view, the means by which the bystander intervenes in order to save the five in BOF—namely, pushing the large man to his death—directly infringes upon the large man's right not to be killed. On the other hand, the means by which the bystander intervenes to save the five in BAS—namely, hitting a switch and diverting the trolley—does not, in and of itself, amount to an infringement of the lone workman's rights. Whereas pushing people is a violation of their rights even if no real harm comes to them as a result, merely diverting a train does not, in and of it-self, directly violate anyone's rights.³

Thomson declares that this difference purportedly explains why it is permissible for the bystander to intervene in BAS. Because the bystander in this case can maximize utility by saving the five without directly infringing upon the rights of the lone workmen, her doing so is permissible (but not obligatory). Conversely, because the bystander in BOF would directly violate the rights of the large man, sacrificing him in the name of the greater good is impermissible. But even if we agree with Thomson that in BAS the bystander won't *directly* violate the rights of the lone workman, it is nevertheless clear that she will knowingly *indirectly* violate the lone workman's right not to be killed. This is something that Thomson explicitly admits (CITE). But why should the issue of directness or indirectness be relevant where a person's general right not to be killed is concerned?

In answering this question, I want to entirely shift the vantage point from which we have been viewing BAS and BOF. Rather than looking at the two cases from the perspective of the bystanders—and what it is permissible (or impermissible) for them to do—I want to examine BAS and BOF from the point of view of the people whose lives

³ It is unclear that Thomson is correct in assuming that pushing someone—unlike hitting a switch or diverting a trolley—always amounts to a rights violation. What if I push someone in order to keep them from being struck by a moving vehicle? Have I really violated that person's rights—albeit permissibly? For present purposes, I am going to set this difficulty with Thomson's view aside.

hang in the balance. First, let's reexamine BAS from the point of view of the five people who will be killed if the bystander understandably cannot bring herself to hit the switch. Imagine that one of the five workmen has a gun and it becomes clear that the bystander is not going to be able to bring herself to hit the switch. Would it be permissible for the workman with the gun to shoot and kill the bystander if doing so was the only way of getting her to fall onto the switch? In order to figure out how Thomson might answer this question, it will prove useful to consider three scenarios that she has identified as morally impermissible instances of self-preservation:

1. Substitution of a bystander: A runaway trolley is going to kill you unless you hit the switch and divert the trolley onto a lone workman—thereby saving yourself. (1991: 289)
2. Use of a bystander: A runaway trolley is going to kill you unless you shoot a large man on an overpass so that his body will fall onto the tracks—thereby stopping the train and saving you. (1991: 290)
3. Riding roughshod over a bystander: A runaway trolley is going to kill you and the only escape route is over a bridge that will only hold one person. The bystander who is already on the bridge will fall to her death if you run across the bridge. (1991: 290)

In BAS, it is not the case that the bystander will kill the five workmen if she cannot bring herself to hit the switch—rather, she will merely allow them to die. Hence, if we adopt Thomson's permissibility conditions for self-defense, it is impermissible for the five workmen to kill the bystander in order to save themselves. Indeed, by killing the bystander and diverting the train, they would also thereby kill the lone workman—which would be similarly impermissible on Thomson's view. By her lights, in order for X's killing Y to be a permissible instance of self-defense, Y must minimally have posed an immediate threat to X's life (or well-being). Since neither the innocent bystander at the switch nor the lone workman pose a threat to the lives of the five workmen who will be killed by the trolley, the workmen cannot kill them merely in order to save themselves. To do so would be wrong on two fronts according to Thomson as it would involve both a

morally impermissible *substitution* of a bystander with respect to the lone workman and a morally impermissible *use* of a bystander with respect to the individual at the switch.

As Larry Alexander has correctly pointed out, this is a very odd line of reasoning for Thomson to adopt (1993: 56). After all, while she concludes that it is morally *permissible* for the bystander in BAS to sacrifice the lone workman in order to save the five, she also seems to be committed to the view that it is nevertheless *impermissible* for the five workmen—whose lives are in jeopardy—to sacrifice either the bystander at the switch or the lone workman in order to save *themselves*. It is a strange moral principle indeed that makes it permissible for X to kill Y in defense of Z that does not at the same time make it permissible for Z to kill Y in her own defense. Minimally, it is clear that Thomson needs an explanation for why the bystander at the switch can help the five by killing the one when they are not morally permitted to do the same on their own behalf.

For present purposes, however, I want to set aside this issue in order to raise yet another problem for Thomson's view that to my knowledge has so far escaped notice in the literature. Rather than examining BAS from the point of view of either the bystander or the five workmen, I want to look at it from the perspective of the lone workman instead. Suppose that he has a gun and he suddenly realizes that the bystander has decided to hit the switch. Now imagine that as the bystander reaches for the switch in her effort to divert the trolley in order to save the five, the lone workman has the chance to shoot the bystander to keep her from hitting the switch. Would it be morally permissible for him to do so? Keep in mind that as far as the lone workman is concerned, the so-called "innocent bystander" ceases to be a bystander the moment she decides to intervene—especially given how Thomson conceptualizes what it means to be a bystander.⁴ By reaching for and attempting to hit the switch, knowing all the while that

⁴ On Thomson's view, "a person is a bystander relative to a particular situation. Suppose Y is in no way causally involved in X's being at risk of death; that seems intuitively to be a sufficient

doing so will bring about the death of the lone workman, the bystander decides to implement a plan that will harm the lone workman and indirectly violate his general right to not to be killed.

But as soon as the bystander decides to enter the causal fray, she ceases to be a bystander according to Thomson's definition. And while the not-so-bystander may become a hero of sorts to the five workmen by hitting the switch, she will at the same time become an aggressor with respect to the lone workman. So, why wouldn't it be morally permissible for the latter to defend himself against a would-be aggressor who threatens to sacrifice his life against his wishes? Surely, the lone workman in this case is not morally obligated to allow himself to be sacrificed in the name of the greater good regardless of how admirable it would be if he chose to do so. Moreover, if the lone workman chooses to defend himself with lethal force against the not-so-bystander, his doing so would be entirely justified and not merely excusable.

This is a conclusion that Thomson must also seemingly accept given what she has said elsewhere concerning the permissibility conditions of self-defense.⁵ Compare, for instance, the following three cases:

1. Villainous Aggressor: A driver of a truck is knowingly, intentionally, and maliciously, trying to run you over. You can stop him only by blowing up the truck. (1991: 283)
2. Innocent Aggressor: The same as Villainous Aggressor except that the driver is not culpable—i.e., he is not knowingly, intentionally, or maliciously trying to run you over. (1991: 284)
3. Innocent Threat: A large man falls accidentally from above. If you shift the position of your awning, you can deflect him but he will be killed. If you do not shift the awning, he will fall on you and kill you, but he will not die. (1991: 287)

Thomson believes that if you happen to find yourself in a situation such that if you do not kill someone else they will bring about your death instead, you are justified in killing

condition for Y's being a bystander to the situation that consists in X's being at risk of death" (1991: 298).

⁵ On her controversial view, self-defense is permissible even in cases where the threatening party is entirely guilt free with respect to the threat (1991: 286).

them first regardless of whether they are trying to kill you knowingly, intentionally, or maliciously. As she says, “once we agree that he is about to violate your right—and that you can prevent this only by killing him—it seems right to conclude that he no longer has a right that you not kill him” (1991: 301).

This remark reveals two important features of Thomson’s overall view. On the one hand, it shows that she believes that whenever someone threatens to violate your general right not to be killed, he thereby gives up his own right not to be killed—regardless of whether he is at fault for the violation. On the other hand, the very same potential rights violation that justifies the act of self defense in the first place also undercuts the violator’s right to defend himself in return. For instance, if you impermissibly attack me and I permissibly defend myself in return, you can’t then use self-defense as grounds for attacking me again. You gave up your right to self-defense when you threatened my life in the first place.

But if we adopt this account of self-defense, then the lone workman would have the right to defend himself against the bystander. Moreover, the bystander would not have the right to defend herself in return. After all, the bystander in BAS surely falls somewhere between being a villainous aggressor and being an innocent aggressor. For while she is not acting with malicious intent, she nevertheless knowingly decides to implement a plan that will (indirectly) infringe upon the lone workman’s general right not to be killed in the name of the greater good. As we have already seen, the lone workman is not obligated to defend himself under these conditions, but he would certainly be well within his right if he chose to do so.

In this respect, BAS is much more similar to BOF than Thomson and others have recognized. Keep in mind that if the large man turned around just in time to realize what the bystander was intending to do—namely, send him plummeting to his death in order to save the five workmen—he, too, would surely be justified in killing the bystander in

self-defense if doing so was the only way to prevent being pushed over the rail himself. If what I have said so far is roughly correct, then the lone workman and the large man have the same status with respect to self-defense. In both BAS and BOF, in deciding to inject herself into the situation, the bystander gives up her causal and moral neutrality with respect to the outcome of the situation. Consequently, both the lone workman and the large man on the footbridge will be justified if they use lethal force in defending themselves against the not-so-bystander. But this just reveals the tension that exists between what Thomson says about BAS and what she says elsewhere about self-defense.

Keep in mind that on Thomson's view X is justified in killing Y in self-defense if and only if Y will (impermissibly) violate X's general right to life. If what I said earlier is correct, then the lone workman in BAS would be justified if he shot and killed the bystander at the switch if it became clear that she was going to divert the trolley onto the side track by hitting the switch. But if justified self-defense requires a potential impermissible rights violation and the lone workman is justified in defending himself from the decisions and actions of the bystander, then it follows that the bystander's decision to sacrifice the lone workman is impermissible after all. As it stands, anyone who wants to defend the intuition that it is permissible for the bystander to hit the switch in BAS owes us an account of self-defense that is consistent with this intuition.⁶

Bibliography:

- Alexander, L. 1993. "Self-Defense, Justification, and Excuse." *Philosophy and Public Affairs*, Vol. 22 (1), 53-66.
- Thomson, J.J. 1985. "The Trolley Problem." *The Yale Law Journal*, Vol. 94 (6), 1395-1415.
- _____. 1991. "Self-Defense." *Philosophy and Public Affairs*, Vol. 20 (4), 283-310.

⁶ I would like to thank Susan Feldman, Adam Feltz, Alfred Mele, Eddy Nahmias, and Jeffrey Reese, for helpful comments on earlier drafts of this paper.

