Haberdashers' School

Occasional Papers Series

Occasional Paper Number Ninety-Three

Reconstructing the Trolley Problem

Sela Arnsberg

Department of English, Haberdashers' School

sarnsberg@habselstree.org.uk

January 2026

Haberdashers' School Occasional Paper Number Ninety-Three
January 2026

Reconstructing the Trolley Problem

Sela Arnsberg[1]

Abstract

This essay will investigate the trolley problem by inserting extra context for the agent involved to see what insights this will reveal. The essay will examine the ways in which one can critique philosophical thought experiments and arguments in order to show the issues with current philosophical practice. This in turn will better explain the methodology and results that are yielded by the augmented trolley problem. The essay will show that one should always find a third option, and that these options can only be found if one has worked hard at self-improvement before the situation arises in accordance with Aristotelian virtues and deontological actions. Further, this essay also finds that, when backed into a corner where such a third option cannot be found, one should do what is best utilitarianly for the agent's chosen ideal. The essay does not offer an ideal for one to strive towards but shows that striving for an ideal allows one to act more in accordance with that ideal's morality.

You are in the driving carriage of a trolley going down the tracks. The brakes are broken, so slowing down/stopping is impossible. However, you can change tracks. This is great news, seeing as there are five people tied to the tracks in front of you. But just as you are about to change the tracks and go down another one, you realise there's a problem: there's *another* person tied down on the other tracks. There are no other tracks, the trolley is barrelling down about to run over five people.

[1] I am indebted to the great Sailee Khurjekar for her brilliant comments that helped shape this essay.

For the sake of space and time, we shall assume that all people tied to the tracks are of the same moral bearing. As such, it is not, for example, the lives of five racist, serial killers that have expressed the desire to kill again given the chance, against the life of an innocent four-year-old. We assume that these are all adults with no dependents and all doing a generally similar and ethically neutral job for an ethically neutral company. If so, and with the assumption above which means each person's life, on any metric, is equal ... what do you do?

In other words: the trolley problem. Again. It seems inevitable that certain problems will never leave, and the thought experiments that they sire along with them. Shame, really, because you would hope that a species that prides itself on its ability to think would have solved something at some stage. Moral problems, however, do not seem to be going anywhere. In fact, they seem too often to be going over the same issues without an end in sight – I would go so far as to say that it is almost a rite of passage when it comes to philosophy that one takes a stab at these topics.[1] Having said that, and in good and proper academic fashion, I shall be doing the same. The trolley problem has not ceased to amaze, annoy and, in the right hands, create a good deal of entertaining television. This essay, instead of fully trying to solve the problem, will try and offer what is missing from the picture – the overarching situation. I will begin by doing something somewhat unphilosophical and show why we should probably pay more attention to literary geniuses and then merge some such insight into the blue philosophical vein.

[1] Usually in an undergraduate essay which will be marked by someone who does not, really, want you to succeed beyond what they have already published in their own right (which they usually have a copy of to disperse, free of charge!) and then, once marked, chucked into the heap of essays that no one looks at until they, in turn, become academics and so do the same thing to the new generation.

To begin with the unphilosophical, I shall first attempt to show the stupidity of the thought experiment. This, of course, is considered gross misconduct in the philosophical world[1] for two reasons. Firstly, academics need to make a living and sound clever, so we need to let them have their say. This tradition follows Socrates very closely, who, in Plato's many writings, always seems to convince his interlocutors that his thought experiments and their conclusions are obviously correct. Seeing as they are usually rich young men who are spending too much time with their local influencer, I think this is sadly all too realistic. Secondly, it defeats the purpose of all thought experiments as one knows that any thought experiment, like any metaphor or scientific experiment, is flawed – they require one to take a few things as accepted and assumed in order to function. This is seen very clearly in the trolley problem: who manages to place five people on one track, tied up, and one person on the other track, also tied up, all while also making sure that the brakes no longer work in the trolley? This is absurd, unless one introduces a chaotic force, such as Stan Lee in the role of God, which would mean that Green Goblin has once more played a trick on Spiderman (probably involving Mary-Jane while he's at it), which is not part of this reality.

Philosophically speaking, the absurd situation concocted cannot give us any real and usable insights into moral theory. Similar to Hume and other's criticism of 'science in a box', where we cannot expect a sterile scientific experiment to provide us with insights into a complex and interconnected universe, these thought experiments that remove all context and agency of an individual ludicrously reducing the choices

[1] But actually seen as very good practice in literature, especially if one is brave enough to go to those witch trials known as 'writing workshops'... you have been warned!

we can make to a simple binary, cannot elicit any insight or deductions into moral actions.

But, we are not allowed to point out such illogical aspects of the thought experiment; we create reality-adjacent experiments to show our intuitions or logical processes to better exemplify the issues/solutions of the problem, which usually happen to support the academic's views, but we shall let them have that. As such, one should not point out the impossible nature of the thought experiment, as it does not allow for the obvious, built-in fallibility as no one situation can be the same as another, but we must try and investigate these ideas somehow!

This is all to say that I am aware that most philosophers will not like what follows, but I do believe that, not only do they need to hear it, but that it will allow us to better understand some of the ideas inherent within the trolley problem.

So, to continue with the example: what do you do? Well, many have argued about the different permissibility of the actions and ways of defining these actions,[1] so I will leave them to their discussions, similar to the way a toddler must bang a toy train into a wall over and over again in order to understand that, unless they are a metahuman, that wall is going to remain. All the solutions, as far as I know,[2] have assumed that the solution has two options: pull the lever or let it be. But I disagree, which is where the philosophical heresy begins.

In philosophy, usually, there are two ways of critiquing the assertions of the philosophers: either you demonstrate that the final conclusion does not follow from

[1] Judith Jarvis Thomson's 'Killing, Letting Die, and the Trolley Problem' (*The Monist* 59 (1976) pp. 204-217) is an excellent paper which tackles the different permutations of the problem and the different theories that have arisen to engage with it.

[2] And I definitely would like to be refuted here – it would be nice to have other members of 'The Thinking Club'.

their process, or you demonstrate that one of the steps in their logic is wrong, but you must never critique the starting example. However, I refuse to play by these rules. After all, why would I not try to flip the trolley, or shut down the engine in some way? One could flip the switch at the right time, which would lead to the trolley derailing and thus all are saved. This, of course, is the literature academic's response, as it shows the plot holes in the story, namely that the character is forced into a binary situation when in fact life does not function that way. The philosopher would respond that this misses the point, but I believe that that response highlights the issue! The fact that we immediately assume that there are two options points to our moral failings! I call this the third-option doctrine.[1] The idea is very simple: we should always search, as a matter of moral principle, for a third option when neither option is a morally acceptable one. Now, many would argue that it is morally right to kill the single person, while some argue that we should let the five die (philosophers are weird), but none, as far as I can tell, would say that if a third option would present itself which saves everyone we should ignore and kill the people tied to the tracks.[2]

The assumption of a lack of choice speaks to the philosopher's inability to think beyond the seeming situation. So: we should always try and find a third option when the two available to us are unacceptable, rather than compromise our principles/morals, whatever they may be. Thus, we must surmise, the lack of an option is a moral failing in and of itself. But how can this be? Surely, being unable to see a solution due to the context of the situation is not a moral failing, as one has yet to do an action? This is, in some sense, true, but it misses the overarching argument: the

[1] When I wish to sound fancy and like a philosopher, usually when I am wearing tweed and with a nice glass of whiskey. Otherwise, I call it the Superman Principle, seeing as the character embodies precisely this ideal (in most of the works he features in, anyway).
[2] In fact, this is part of the Double Effect doctrine, which I will not explain here – look it up!

person has failed in their moral duties when they have not prepared sufficiently for such a situation. At first glance, this seems absurd. How can we be expected to prepare for every situation? That is impossible and unreasonable. But it is only so if we assumed that that is the requirement. To deal with as many situations as possible, to best be able to find that third option, we must endeavour to constantly better ourselves and study. We must learn both concrete knowledge and its limitations, to better our abilities, both cognitive and physical, so that we stand the best chance of thinking up a third option. Thus, in the trolley problem, if we only see the two options, the moral failing has already happened, as we have not prepared ourselves to think creatively about situations, whatever they may be. We might not have exercised well enough to make sure that we are physically able to handle the stress or operate the machine. Thus, the moral failing has happened before we even set foot on the trolley, and we need to accept that responsibility, hard as it may be to do so.

To put this in philosophical frameworks, we should use Aristotelian-like virtues in accordance with our ideals, enforced and strengthened by habit and learning, to act deontologically when difficult situations arise, whence we can find that third option, which is a two-fold deontological action: it is a morally right action to find a third option in and of itself, and then it becomes our moral deontological duty to carry out that third option. Note that, at this stage, we are not considering consequentialism of any of kind. This is because if we have done well before the situation and during the situation, then the best consequence will happen by itself, without us having to consider the consequences. Arguably, we are thinking consequentially when considering why we habitually practice what is a virtue, but this is misleading: self-improvement is a good thing in and of itself, both as an ideal and as a deontological duty when enacting it –

the consequence of it may never be known, seeing as the odds of Green Goblin forcing us into the trolley problem are essentially nil.

Morality does not stem from the choices alone, but from the person making those decisions. After all, we do not say that a three-year-old is immoral because they punched another or snatched a toy! At most, they could be said to be amoral, but how can we even say that, really, if they are in the process of understanding what rudimentary morality is? We cannot attribute a claim of such magnitude to someone that has had no chance of earning it. Thus, despite the wringing of philosophical hands at the horrid blasphemy of messing up their perfect thought experiment, I think we have exposed a key fact regarding their morality: they expect us to be simple-minded and unempathetic, thus making all our lives constantly about ourselves, never thinking creatively, never building our empathy for others to the point that we want to better ourselves in order to help them in as many situations as possible.

However, this is not the end of the story. While we have seen that we should search for the third option, this may not always be possible for two reasons: first, as shown above, we have already failed in our moral duty to gain the skills, knowledge and ability to find and execute that coveted third option, and must therefore choose one of the options before us. Second, there could be some constraints. For the sake of this argument, let us suppose that the evil person who tied the people to the tracks has informed us (and we have verified this in some manner) that the switch will not respond after a certain point down the tracks, nor that we can do anything to the engine, etc. Thus, both the person that has failed their duty and the one that has not are in the same position and must decide. What do you do?

The answer comes from the same position: a derivative of the Hope principle, which is at the core of the action, the ideal through which one who would act as I suggested above must strive. In the continued circumstances of the trolley problem in which we find ourselves, and using the theory I am currently espousing, this means that we must do that which gives the greatest amount of Hope. As far as I have read, most arguments do not consider what the ideal that is driving the decision is, but rather what can be said of the decision outside of any context, either situational or ethical. Thus, by inserting a moral ideal that we can strive towards, a moral ideal that we can hope to enhance with our actions, we can find a consistent guide to our actions.

In this case, we must kill the one person to save the five. Seeing as all six people are of categorically equal moral worth – as assumed at the start of the essay – we must do that which is most Hopeful. The act may be horrid, but the act is also to save five people, to allow five people to continue (hopefully) to better themselves and thus save others in time. In such situations where we cannot save everyone, we must save as many people as we can.[1] Somewhat contradictorily, this would mean resorting to utilitarian principles. But that is what happens when all else fails: when the habits we have trained fail us and no action can be said to be deontologically sound then we must resort to utilitarian principles that would support the ideal we strive towards.

At least, that's what I think. I won't argue the merits of the ideal here, simply because it is secondary to my main point: that if we insert an ideal through which we act, by which we are guided in our day to day life, we find a consistent way of dealing with

[1] This is still assuming morally equal people, there could be circumstances that may make us choose differently.

situations.[1] This, however, would suggest a form of moral relativism, as it would suggest that the acts would change depending on the circumstances, so long as some vague consistent ideal supports the action(s). It is not my aim here to argue for any specific principle, but I would say that any principle that does not have strong, compelling arguments for acts that should or should not be done is liable to be a morally weak one, which should be replaced by something else.

Before we finish, we have one more contextual factor to think of, which, again, is not referred to: guilt. We assume that we should, if we manage to find a strong answer to the trolley problem, not feel any guilt: we did the right thing, what else is there to do? Surely, one would argue, we have nothing to feel guilty about. This is a crude way of thinking, indicative of far too many philosophers that do nothing in life and therefore feel that these issues of guilt are irrelevant, as they have nothing to be guilty of, having done so little. The kind of philosophers that, when the revolution they espoused for so long happens outside their classroom, sneak home and write an essay about it rather than being with the masses they inspired.[2]

We should feel guilty whatever we chose to do for the ideal we follow and the answer we came up with. We should feel guilty. Maybe we could have unhacked the trolley? Maybe we could have worked harder before so that the third option would have been viable? Whatever the case, we should feel guilty. But though we should *feel* guilty, that does not mean that we are necessarily guilty. The one that put them on the tracks, not us, is the guilty one. In the permutations where they have fallen to the tracks from above and the sides are too steep for climbing it is the people on the

[1] To choose a horrid and preposterous extra example to show the main argument *(not* my position): if the ideal were the elimination of humanity, then one should have found a way to kill all six of the people on the tracks. But, failing that, kill the five – the doctrine can still work with any principles or ideals.
[2] It is important to note that not *all* philosophers are like that.

tracks and the ones that designed the train tracks area so poorly without any safety space. I am not saying that the planners needed to account for this precise eventuality, but they should have thought of the need for enough space for someone to step off the train if it had to be stopped at any point, at the very least.

What I am trying to say is that the guilt in this situation does not solely fall upon the one at the control panel. It is a whole network of people who are guilty and we need to acknowledge that. We need to accept the difficult fact that not doing what is right can lead to others' harm, and that this harm may occur and pass without us even knowing it happened, nor be aware of the complicity we have in the event. We can create different circumstances for the people on the tracks 'til the cows come home, but that is not my objective here. The objective of this essay is to show that in narrowing the scope of the trolley problem and its derivatives in the manner often done provides us with no real insights into human behaviour or our intuitions, but simply allows armchair-ridden philosophers to experience difficult situations without actually having any stake in the situation. It is only by taking a person as a person, with ideals that they strive to emulate through their actions and considering the context of the situation, that we can arrive at any meaningful answer to the trolley problem.

So, what is the solution? Simply: to each their own. A cop-out? Not really. It is a call to arms, an invitation, the harbinger of your own moral development. You, yes YOU, need to decide what you believe in, work hard to live by that code, even when it is difficult. When that code no longer functions, you must augment it or adopt a better one; question it every day with brilliant thought experiments like the trolley problem - see if you still agree with the principle! I find that, by and large, most people want to

do good, and that that good, by and large, is a universal good. So, I believe in your ideal, and I believe in what you will end up doing in such situations – though I do hope that it never happens.