

David Edmonds on Trolley Problems

David Edmonds: *Is it ever acceptable to take one life to save five? To discuss this topic, we have an unusual guest who has written about this. You may recognise his voice.*

Nigel Warburton: *David Edmonds, welcome to Philosophy Bites.*

DE: Good to be here. I've heard great things about the show.

NW: *The topic we're going to focus on is "The Trolley Problem". Could you begin by outlining what the trolley problem is?*

DE: Well, there are lots of trolley problems. In fact, they've been given a name: trolleyology. These are all thought experiments. And the first and most basic one is this – you imagine that a train is running out of control (some talk about trolleys, but in fact it's easier to talk about trains). The train is running out of control; it's going down the track, the brakes have failed and on the track – ahead of the train – five people are tied. They're going to be killed unless something is done. You're standing by the side of the track. You can turn a lever which will turn the train down the side track. Unfortunately, on that side track, one person is tied to the track. The question is: should you turn the lever and so turn the train onto the side track to kill one person to save five? What's interesting about that problem is that, if you ask people what they should do, pretty much everybody thinks you should turn the train and kill the one to save the five.

NW: *So, that's a straightforward case. It's difficult because you'd have to act quickly. You might feel psychologically traumatised by your action but you would be doing the right thing, most people would think. If you flick the lever, the train's diverted, it kills one person instead of five.*

DE: That's right. Now, there's another trolley problem known as "The Fat Man Problem." This – again – involves a train running out of control, about to kill five people. This time, you're standing on a footbridge. You're standing next to a very fat man. If you were to push the fat man over the footbridge, the fat man would splatter down on the track below and his bulk would stop the train in its track and therefore save the five people. But the fat man would die. It's important that the man is fat because, otherwise, the moral thing to do would, obviously, be to jump over the footbridge and sacrifice yourself to save the five. But you can't do that because you're not fat enough to stop the train, so your only option is to push the fat man over.

NW: *So this thought experiment only works if you're slimmer than the fat man.*

DE: Exactly. And if you ask people what you should do in that case, pretty much everybody believes that it's wrong to push the fat man. So this is the heart of the trolley problem. In both trolley examples that we've set up, you have the choice between allowing five to die and killing one. In the first case, almost everyone thinks it's right to turn the train to kill the one and to save the five. But, in the second case, almost nobody believes that it's right to push the fat man to save the five.

NW: *Yet, from a consequentialist point of view – somebody who assesses the moral worth of actions by their probable consequences – it looks as if the two cases are similar. Your actions result in five*

people living who wouldn't otherwise have lived and one person dying who wouldn't otherwise have died.

DE: That's exactly the puzzle. And that's a puzzle that has been with the world of philosophy for half a century now. It was first thought up by a woman called Philippa Foot, who was an Oxford-based philosopher. And it was then taken up by an American philosopher called Judith Jarvis Thompson who gave us the fat man. And it remains a problem today. What is the explanation for the difference in our intuitions? And it's been like a philosophical crime mystery – really – or a logical puzzle that philosophers have been trying to solve for many years.

NW: *Well, one psychological explanation may be: it's pretty easy to pull a lever but it's quite terrifying to wrestle with a large person – a person larger than you, in a life or death struggle. And it's hopefully, from your point of view, going to result in that person's death.*

DE: Right. So what you're offering there is not a *philosophical* explanation, you're offering a psychological explanation. You're saying we're squeamish about the one and we're not squeamish about the other. Well, that's quite plausible. But imagine another thought experiment. The train is going along and it's going to kill the five. Once again, the fat man is on the footbridge. This time, you're not standing next to him; you don't have to push him over. You're exactly where you were in the first scenario by the side of the track. This time – again – you can turn a switch – a simple little arm movement – but what happens now is that a trapdoor opens and the fat man comes crashing through and lands onto the track, in front of the train, and he dies and so saves the five.

NW: *How do people respond to the refined fat man case with the trap door and the lever?*

DE: That's the interesting thing. If you ask them that case, they continue to think that it's wrong to kill the fat man, even though this time they're not having to push the fat man. So there's no assault, there's no battery; all they're doing is turning a switch just as they were in the first case. And yet, most people continue to think that it's wrong to kill the fat man. Not as many think it's as wrong as when they actually have to physically push the fat man, but still the majority. So there's much more going on than the mere squeamishness of having to engage in battery and physical assault.

NW: *This is really interesting. So the psychological explanation can't explain everything because some people – many people, in fact – are still quite reluctant to pull the lever that will release the fat man and stop the train.*

DE: This is the holy grail of trolleyology, trying to come up with an answer to that particular conundrum. And I think the answer lies in what's called the "Doctrine of Double Effect" which is a doctrine that was first invented – or discovered – by Thomas Aquinas a thousand years ago. This doctrine is useful in lots of cases in the real world, most famously in the case of war. So, you often hear politicians condemning terrorists for intentionally killing civilians. Whereas, when an American or Brit orders an operation that kills civilians, they always call it collateral damage. A terrible euphemism but what they mean – what they think has moral significance – is that in those cases they might foresee, in a particular military operation, that some civilians might die in a military operation, but that's not their *intention*. They don't intend to kill these civilians. It's a foreseen consequence but not an *intended* consequence. And I think the doctrine of double effect can be used to explain the difference in our intuitions between the various trolley cases.

NW: *So that's the doctrine of double effect – and it's not uncontroversial – but, let's apply it to the fat man case and the spur case and trolley problems and show how that works.*

DE: To go back to the original spur case, the train is going going to kill five. You can turn a switch and turn it onto the side-track where one person is tied. Do you intend to kill that one person? Well, imagine what would happen in a possible world. The train goes down the side-track and miraculously, the one person on the side-track is able to extricate himself from his ropes and run away. The train would then continue down its path and it wouldn't kill anybody. What would you think if you're the bystander and you've turned the train? You'd be delighted! You've saved the five; the train is no longer heading towards the five and you haven't had to kill the one. Now think about the fat man case, you push the fat man off the footbridge. Imagine the fat man is wearing a big rubber suit and bounces off the track and runs away. Would you be delighted? No! The whole point of pushing the fat man is that he needs to get in the way of the train. You intend his death, as it were. If he's not there, the train will trundle along and kill the five. So, it looks as if there's a difference in intention between the spur case and the fat man case. You're using the fat man to save the five; you need the fat man's death. On the other hand, you don't need the death of the one person on the spur to save the five people.

NW: *But if the fat man managed to stop the train and survive, that would ok. I'd be quite pleased with that. If he was wearing a big rubber suit and the train bounced off him – that would be cool.*

DE: That's a very good objection to what I've just said: and it's linked to the problem or objection of "closeness". Let me give you another thought experiment: imagine that five people are trapped in a cave. The waters are rising; the reason that they can't escape the cave is because a fat man is stuck in a hole in the cave. His head is out, so he can breathe, but the water is slowly rising and the five others in the cave will all die unless – somehow – you can get the fat man out of the hole. You have a stick of dynamite. There's one thing you can do; you can blow up the fat man and all five people can then escape. Now, suppose you use this dynamite: what do you intend to do when you blow up the fat man? Do you intend his death? Well, you could say: no – all I intend to do is to blow him up into a thousand pieces so that the five people can escape. I don't want him to die! If miraculously those one thousand pieces could be stitched back together and the fat man could survive, I'd be delighted! But, the problem is – blowing someone up into a thousand different pieces and killing them are so closely related, it makes a kind of nonsense to say that you want to do one without the other.

NW: *But to me, there's a bigger problem here because the doctrine of double effect doesn't seem that plausible. We might make fine grain distinctions about whether someone is more or less culpable but generally, it's the consequences that matter. We should want to save five people. So if someone was squeamish and didn't kill the fat man, you might want to say that they were culpable for not saving the five.*

DE: I think the doctrine of double effect can be open to abuse. You can imagine politicians saying "Oh, I didn't intend to kill the civilians; it was merely a foreseen effect, it's not as bad as if I intended to kill them." The question for philosophers is not how this concept can be *abused* but whether it has any philosophical validity. One way to think about it philosophically is as follows: imagine you have a politician who says "I didn't intend to kill the civilians, I merely intended to destroy the military installation" and they've sent a couple of missiles over to destroy the military installation. Imagine the civilians now run away. If the politician really did intend to kill them, they'd send

missiles to where the civilians have now moved. So by imagining counterfactual scenarios, you can put flesh on the difference between *intending* something and merely foreseeing it. And I think this distinction has moral salience.

NW: *Isn't this a distinction that is used often by religious people who will say something like it's ok to administer pain-relieving drugs – opiates, let's say – to someone who's in great pain, with the foreseeable consequence that it will shorten that person's life. But it would be absolutely wrong to deliberately commit an act of euthanasia?*

DE: It is used by religious people. I've mentioned this begins with Thomas Aquinas, and it's become a central part of Catholic theology. But it's now embraced a lot more widely than just by Christianity; it's accepted by many secular philosophers, as well. But you're right. It's a doctrine that is used in many cases in the real world, euthanasia being one example, abortion another example where some people say it's wrong intentionally to kill the foetus but it's not wrong sometimes to kill the foetus merely as a side effect of saving the mother's life. And, we've mentioned, war's another classic example where the doctrine of double effect is cited routinely.

NW: *The trolley problem's been around for quite a long time. It was first formulated in the 60s, got a new lease of life in the 80s. And, it's back now in several different ways. One of the ways is that some people are talking about it in relation to what's natural and universal in human beings – that we have some innate moral grammar.*

DE: That's right. A parallel has been drawn between Chomsky's work in linguistics and the work of people such as John Mikhail in philosophy. And philosophers like John Mikhail argue that we probably do have a *universal* moral grammar and trolley problems provide some evidence for this. And there've been tests around the world; tests involving men and women, the old and the young, the rich and the poor, people who live in cities and people who live in the countryside. And the remarkable thing is that intuitions are roughly the same wherever you are and whether you're eight years old or eighty years old – and that does seem to indicate that there is some kind of universal moral grammar.

NW: *Another new boost to discussion of trolley problems has come from neuroscience.*

DE: Neuroscientists have taken scans of our brains whilst subjects are thinking about these moral problems. And they've worked out which bits of the brain, –very, very crudely put –light up when we're pushing the fat man or when we're turning the train. The really interesting question for philosophy is what you do with this stuff. We're finding out all this factual information about how the brain works morally; what happens if a bit of the brain is damaged and whether subsequently we're more likely to come to utilitarian judgements as a result of an accident – say – or a lesion. But what are the implications of these findings? Just because we know about some of the brain mechanisms at play, doesn't tell us – or does it? And this is the key question – what we should *do* in moral dilemmas. There's some evidence that the *calculating* part of the brain is the bit of the brain that does the work in the original spur case and the *emotional* part of the brain is what is holding you back from pushing the fat man over the footbridge. Now, one very eminent philosopher, Peter Singer – who's perhaps the world's most famous living utilitarian – has argued that we should privilege our analytic/calculating side and that the neuro-evidence is evidence for us having made a mistake in our reluctance to push the fat man. But we could argue quite the opposite. People who

have got damage to their emotional part of their brain are more willing to push the fat man. And we might want to say, well that's evidence that pushing the fat man is *wrong*. After all, these are damaged patients. So I think, over the next few years, this is going to be the most interesting part of moral philosophy – to see how the empirical work finds its way into normative questions about how we should behave.

NW: *Aren't these trolley problems a bit like chess problems? They're interesting, perhaps there's a solution and you want to solve them but they don't really impact real life.*

DE: Well, they are fun but that's not necessarily inconsistent with there being useful. The great advantage of these trolley problems is that real life cases are full of background noise – white noise – there are many things going on. They're very complex and it's not quite clear what's doing the work in shaping our intuitions and the great thing about the trolley problem is that you clear the white noise out. And with these pure trolley problems you can begin to manipulate intuitions by changing one variable or another - and in that way you can see what's really going on.

NW: *But does it really have any practical application? Once you've tweaked your thought experiment and come up with the conclusions and worked out the principles, how do you get back to real life?*

DE: The idea is that once you find your principle or principles that seem to explain these different trolley problems, you can then extract them from the thought experiment and apply them to the real world. You can say: well, maybe the doctrine of double effect is something that we should take seriously. And if it's something that we should take seriously, this might help when it comes to thinking about what we should do in certain circumstances in a hospital, for example. Or, what we should do in warfare. We've identified a principle that we think might be helpful in the trolley problems and we can transplant that principle back into the real world.

NW: *So, could you give an example of a real life case which was – in effect – a trolley problem?*

DE: Well, one interesting example is from World War Two. There was a dispute in the British cabinet between one cabinet minister and Churchill. And needless to say, Churchill won – as he always did in these disputes. The V1 missiles were being fired from the continent and they were landing to the south of London, hitting less densely populated areas. They were intended to hit the centre of town. And Britain had a number of double-agents. And the question was, whether these double-agents should be used to send back false information to the Nazis. This would trick the Nazis into believing they were hitting the centre of town or – indeed – hitting north of London with the hope that they would then adjust the trajectory so that the missiles would fall even fewer south and kill fewer people. Well, the cabinet minister who was opposed to Churchill said "We can't do that, we can't play God! If we do that, other people will be killed; we'll have blood on our hands, we'll be responsible for that". Churchill, in that case, took the very utilitarian decision that actually that was exactly what we should do – that we should be trying to minimise the loss of life. And this is a very good parallel to the first trolley case: the train that's going to kill five unless we divert it to kill the one. So, in the trolley problem, Churchill shared the intuition of most of the rest of us – that in this circumstance it is acceptable to kill one person in order to save five lives.

NW: *David Edmonds, thank you very much.*

DE: Thank you Nigel, I've enjoyed it. I hope you'll have me back one day. [ends]