5 RIGHTNESS, BLAMEWORTHINESS, AND THE DOCTRINE OF DOUBLE EFFECT

Caroline Lyster,[1] *McGill University*

According to consequentialist moral theories, the right thing to do is whatever brings about the best consequences. While this does seem to be the way that we make decisions in the real world, a strict application of consequentialist reasoning can sometimes lead to our allowing, or even doing, harm to others in order to bring about good. This seems intuitively wrong. We need action guiding principles to prevent this from happening; principles that will require us to help people and, at the same time, prohibit us from harming them. But, a strict application of principles like beneficence, or non-malfeasance, can also lead to counterintuitive results in cases where the seemingly 'right thing to do' results in collateral damage. What is going on here? Either, we must further modify our action guiding principles, or our intuitions about these collateral damage cases are wrong. I believe it is the latter; in these cases we are confusing intuitions about what is right with intuitions about what is not blameworthy. This means that the doctrine of double effect, the principle that is invoked to explain the apparent permissibility of collateral damage, is not a deontic one, but an aretaic one.[2]

In order to understand where the doctrine of double effect comes from, we should first understand what it is invoked in response to. Con-

1 Caroline Lyster graduated from the *University of Alberta*, in Edmonton, Canada. She is currently reading for a Masters in Bioethics, at *McGill University*.

2 The deontic principle points toward what is right, and the aretaic tells us whether or not a person should be blamed for what they have done.

sequential moral theories are concerned with outcomes, and the right thing to do is whatever leads to the best consequences. On these theories actions have only instrumental value; that is, they are not good or bad in themselves, they are only good or bad insofar as they bring about good or bad consequences. The answers given by consequentialism are usually in line with our intuitions about cases where moral decisions must be made. Consider the following example:

> Transplant 1: A doctor has access to five organs for transplant and has two choices for how to distribute them. He can (1) transplant all of them into one patient who is suffering from the failure of multiple organ systems or (2) transplant each organ into one of five people who are all suffering from failure of a single organ.

According to consequentialism, the right action would be the second option, where the organs are used to save the maximum number of people, and therefore, do the maximum amount of good. Most people would intuitively agree that this is the right choice.

But, if it is all about the consequences, what is the right action in this second transplant case?

> Transplant 2: A doctor has five patients, each dying from the failure of a single organ. One day, a healthy young person comes into the clinic for a simple check-up. As it turns out, this healthy young person is a match for all five of the doctor's dying patients.

If actions are right or wrong only insofar as they bring about good or bad consequences, then the right action in this second transplant case should be to kill the healthy patient and distribute their organs to the dying five. However, this seems intuitively wrong.

Our intuitions about the second transplant case demonstrate the need for forward-looking, or deontic, principles that can judge certain actions as right or wrong. Since most consequentialists would accept some version of the principle of beneficence, this seems the logical place to start, so action guiding principle number one should be that we should act in ways that help others. Beneficence would explain why we feel the way we do about the first transplant case: the right action is to distribute the five organs to five different people because this action helps more people. A second deontic principle can be deduced from our intuitions about Transplant 2: obviously the reason it is wrong to kill the healthy patient in order to save five others is because to do so would be to harm them. Therefore, action guiding principle number two may be something like non-malfeasance: do not only act to help others, but also act in such a way as to not harm them.

Non-malfeasance as a guiding principle is certainly a good thing, but we should further define what exactly is ruled out when we say, "do no harm". The most straightforward way of interpreting this statement would be as a prohibition against actively doing harm as opposed to merely allowing it to happen, and this distinction does capture why we feel the way we do about the two transplants: in the first transplant case, the one person is merely allowed to die whereas in the second, the healthy patient must be killed. However, as James Rachels has shown elsewhere (1975), the distinction between doing and allowing harm is tenuous, and ultimately seems to amount to not much more than a desire to keep our own proverbial hands clean. So, however we formulate it, our principle of non-malfeasance must account for harms that arise from intentional omissions. In other words, we need a deontic principle that accounts for the intuitions we have about situations like this one:

> The Trolley and the Fat Man: You and your tour group are sightseeing in a trolley yard when you spy five people tied to the tracks some way down from the platform you are standing on. Then, you notice that, an out of control trolley is rolling towards them. Among your tour group members is a fat man, and out of the corner of your eye you spy the tour operator sneaking up behind him with the obvious intention of pushing him in front of the trolley to save the five on the tracks.

If the principle in question simply prevented us from harming others, we could let someone else push the fat man in front of the trolley, but our intuitions here seem to indicate otherwise. A better formulation of non-malfeasance would be something like what Warren Quinn proposes in his version of the doctrine of doing and allowing (1989). In here, when we say, "do not harm", we prohibit both positive and negative agency, and actually mean something more like, "do not harm or deliberately fail to prevent harm". This version of the principle prohibits sitting back and letting someone else push the fat man in front of the trolley, while still permitting the non-deliberate harm that occurs in the original transplant case.

These straightforward deontic principles of beneficence and non-malfeasance appear to capture our intuitions about why certain actions are right or wrong, but there are instances where a strict application of them would have counter-intuitive results. These ones, are cases where it seems as though doing the right thing will result in collateral damage. For example:

> Classic Trolley: You are walking along, minding your own business, when you come across five people tied to the trol-

> ley tracks. Then, you notice that, an out of control trolley is careening down towards these five people and will kill them if you do nothing. You spy a switch, and discover that if you were to flip it, the trolley would be diverted onto a sidetrack, saving the five. However, there is a single person tied to this sidetrack, and if you divert the trolley this person will die (Foot, 1978).

What should you do in this case? Most people would say that they should flip the switch, but that action would violate the principle of non-malfeasance, which requires that we act in such a way as to not harm others.

The fact that we have this intuition indicates that our current deontic principle of non-malfeasance is insufficient. We need a way to prevent harm from coming to the healthy patient in Transplant 2, while at the same time allowing for diversion in Classic Trolley. The solution that has often been proposed for this problem is the doctrine of double effect (DDE), which would further refine our basic principle of non-malfeasance. According to this version of the principle, when we say, "do no harm", what we actually mean is: "do not harm or deliberately fail to prevent harm in order to benefit someone else". In other words, the kind of malfeasance that is prohibited by the DDE is that where someone is benefitted at the expense of someone else. This modification to non-malfeasance allows us to take an agent's intentions into consideration; when we use the DDE to evaluate an action, we ask ourselves whether the agent performing it is aiming toward the harm being cause, or whether that harm is merely a side effect of the action that will be done. Using this principle would allow diversion in the Classic Trolley case because the five people are not benefitted at the expense of the one. Rather, the single person's death is a mere side effect of their salvation. The DDE also gets us the results we want in the transplant cases. In the first case, we are allowed to give the organs to five people, and not just one, because that person's dying is not our intention. In Transplant 2 the doctor is not permitted to kill the healthy patient because he would intend that person's death.

While the DDE seems plausible at the out set, it is not without problems. Foremost among them is this: why should an agent's intention be the slightest bit relevant in determining the right thing to do? Saying that intention matters would mean that the right action, would only really be so, if the agent did it for the right reasons. But, if that agent has the wrong intention, what might have been the right choice, would suddenly become the wrong one. To illustrate this point, consider a slightly different version of the classic trolley case:

> Paul and the Trolley: You are walking along, minding your own business, when you come across five people tied to the trolley tracks. Then, you notice that, an out of control trolley is careening down towards these five people and will kill them if you do nothing. You spy a switch, and discover that if you were to flip it, the trolley would be diverted onto a sidetrack, saving the five. However, there is a single person tied to this sidetrack, and if you divert the trolley, this person will die. Upon closer inspection you discover that the single person is your mortal enemy Paul. You have no real desire to save the five, but because you really hate Paul, you decide to flip the switch.

In the Classic Trolley our intuitions indicate flipping the switch is the right thing to do, but as soon as we flip the switch in the new scenario it feels as though we have done something wrong. In some cases, not only can considering intention make the right action wrong, it can also make the wrong action right. For example:

> Mom and the Trolley: You are walking along, minding your own business, when you come across five people tied to the trolley tracks. Then, you notice that, an out of control trolley is careening down towards these five people and will kill them if you do nothing. You spy a switch, and discover that if you were to flip it, the trolley would be diverted onto a sidetrack, saving the five. However, there is a single person tied to this sidetrack, and if you divert the trolley, this person will die. Upon closer inspection, you discover that the single person is your mother. You have nothing against the five, but decide not to flip the switch in order to ensure your mother's survival.

If the DDE is right, and what you should do in a given situation is linked directly to your intentions, then there are three different right actions in each of these trolley cases: (1) it is right to divert as long as you want to save the five, (2) it is wrong to divert if you only want to hurt the one, and (3) right to not intervene if the one is a person whom you want to save. But, can this actually be the way things work? Is it really only right to divert in some cases? Such a notion is inconsistent with the very idea of objective moral rules, and a good deontic principle should be true, regardless of the person performing the action, or the particular situation in which they find themselves.

A second problem with the DDE is this: when is one 'close enough' to intend harm so that it ought to be prohibited by the principle of non-malfeasance? We would like our deontic principle to account for all of

our intuitions about collateral damage, and there are certain intuitively wrong situations that seem to be permitted by the DDE. For example:

> Poison Gas Hospital: You are a doctor working in a hospital. Five of your patients are dying, and the only way to save their lives is to manufacture a certain chemical that will cure their disease. Unfortunately, the gas that is created as a by-product of this synthesis is poisonous, and the lab in which the chemical will be synthesised vents into a neighbouring room. This room is home to a single person who, for some unknown reason, canot be moved (Foot, 1978).

If we acted in accordance with the DDE, we would be allowed to manufacture the chemical to save the five because the death of the single person in the attached room is an unintended side effect. However, our intuitions seem to suggest that this would not be the right choice; there is something about manufacturing the chemical and releasing a poison gas into a neighbouring room that seems awfully close to intending harm. Even though we do not intend to harm the single person, in manufacturing the chemical we set into motion a series of events, which we know will culminate in their death. Perhaps, then, we could say that a person performs an action that is "close enough to intending harm" when they initiate a series of events with full knowledge of the harmful consequences that will result. Given this, our deontic principle of non-malfeasance should be further refined to say something more like this: when we say, "do no harm", we really mean "do not act, or deliberately fail to act, if your doing so will (a) cause someone undue harm, or (b) cause harm to one person in order to benefit another, or (c) set into motion a series of events through which you know someone will be harmed". This version of the principle explains why we may not manufacture chemical to save the lives of the five dying patients in the Poison Gas Hospital case. It also explains why we may not stand back and watch while someone else pushes a fat man in front of the trolley, and why we may not kill one healthy person in order to save five others.

You may have noticed, we turned to the DDE to try and make sense of a single intuition: that of being permitted to divert in Classic Trolley. But, in order to make the DDE work we needed to define exactly what it is prohibiting. Further, we could determine that setting into motion a series of events, which culminate in a foreseen harm, is close enough to intend harm for warrant being prohibited. But, this modification has a strange consequence: it does not allow diversion in the Classic Trolley case. The act of diverting the trolley to save the five sets into motion a series of events resulting in the foreseen death of the single person on the sidetrack. Therefore, this act is prohibited by the principle of non-malfeasance, as it has been articulated.

Let us take stock of the moves that have been made to get us to this point. Firstly, there was consequentialism which, while attractive in it's quest for the most desirable outcome, can lead to problems where agents may be allowed to cause a great deal of harm to one person in order to create good for more. This is intuitively wrong, so there must be a principle of non-malfeasance to balance out the ethical requirement to do good things. In attempting to spell out exactly that this principle of non-malfeasance might amount to, we came to the conclusion that saying, "do no harm" means something more like "do not directly cause or deliberately fail to prevent harm". The problem arose when we realised that this version of our principle of non-malfeasance does not allow for the right action to include collateral damage. We attempted to account for this intuition by invoking the DDE, but quickly found that any useful version of this principle is eventually not going to allow for diversion.

Perhaps, instead of changing our principle of non-malfeasance to account for the Classic Trolley intuition, we should turn to the case and figure out exactly what it is that we are responding to. Logically speaking, if what we are reacting to is the permissibility of diversion, then any case of diversion will yield the same intuitive result. To test this, consider this case, first presented by Peter Unger (1996):

> Trolley on a Hill: You are walking along, minding your own business, when you come across five people tied to the trolley tracks. Then, you notice that, an out of control trolley is careening down towards these five people and will kill them if you do nothing. You spy a switch, and discover that if you were to flip it, the trolley would be diverted onto a sidetrack, saving the five. However, the sidetrack abruptly ends at the top of a hill, and if you divert the trolley, it will end up rolling down the hill and killing an innocent person reading a book in their backyard.

The only difference between this case and that of the Classic Trolley is the location of the innocent bystander: in the latter case, the five are tied to the trolley tracks: however, the single person is sitting in the backyard, some distance away, minding his own business. But, this difference seems to be enough to change our intuitions about diverting. In the Classic case, we want to flip the switch, but in the new case, the answer is not as clear. Is the distance, the diverted trolley must travel to harm the single person, really a relevant factor in determining whether flipping the switch is the right thing to do? It seems arbitrary to say that it is permissible to flip the switch in cases where the bystander will be killed soon, and not in cases where the bystander will be killed later.

Diversion fails the test, so there must be something wrong with our intuitions about the Classic case. We might be able to take a cue from

Peter Unger, who suggests that a mental phenomenon called "projective grouping" can affect the way we think about different cases, and subsequently, how we react to them. According to Unger, "often we view a certain serious problem as being a *problem* for only those folks viewed as bring (grouped together) in a particular situation" (1996: 97). So, the location of the single person in each case makes a difference as to whether or not we group them with the five. In the case of the Trolley on the Hill, the single person sitting in the backyard is not close enough to be counted in the group of people that have a problem. Conversely, in the Classic case, the single person tied to the sidetrack is seen to be as threatened by the runaway trolley as the five. In other words, when we picture the Classic Trolley case, we think of it as looking something like this:

> Two Runaway Trolleys: You are walking along, minding your own business, when you come across two sets of parallel trolley tracks. Five people are tied down to one set, and there is a single person tied down to the other. Then, you notice that, two out of control trolleys are careening down both sets of tracks. You spy a bazooka with only one round, and determine that you can safely blow up one of the trolleys without harming anyone.

This is a case where all six people are clearly in imminent danger, and we may, in this case, be permitted to use our resources to save five rather than one. However, in the Classic case, the single person on the sidetrack is not in any danger until the switch is flipped. We create the danger to him, and given this, our principle of non-malfeasance will not allow us to divert the trolley. But, even if we are careful to keep in mind that the single person, either on the sidetrack, or in the yard, are not in imminent danger, there is still a part of us that wants to flip the switch. Why? I propose that this intuition has nothing to do with what is right, and everything to do with what is blameworthy. That is, we may feel that diversion is okay because we would not blame someone for making that decision.

Whilst deontic principles look forward to tell us the right thing to do, aretaic assessments look back on choices that have been made, and determine whether or not the agent who made the choice is blameworthy. Now, blameworthiness is only somewhat related to whether or not a particular action was right or wrong; when we assign blame, there are other factors that we may take into consideration. One aspect of assigning blameworthiness is the relative wrongness of an action; that is, an agent may due more or less blame depending on exactly what they have done. Consider the example of stealing. It is always wrong, but a person who steals a car may be more blameworthy than someone who steals a

pack of gum, and this relative wrongness plays out in the way that the penalties for stealing are assigned by the justice system. A person stealing a car would be more harshly punished than a person stealing a pack of gum. This difference in 'sentencing' is directly related to how much blame we feel that each person is due.

A second component is responsibility. That is, the more an agent is responsible for their action, the more they are blameworthy. For instance, an agent who kills someone by jumping on them from some height can be blamed more than someone who is pushed and falls. We are also tempted to view people as less blameworthy if they act, and appeared to be right, given the information that they had. Additionally, we do not blame people for failing to do the right thing, when it would be very difficult for them to do so. Unfortunately, none of these aspects can explain our blameworthiness intuitions in the Classic Trolley case, as diverting the trolley, and killing the tied-person, are definitely on the bad side of relative wrongness.[3] But, we still want to save the five, so there must be another reason why the choice is not blameworthy.

I believe that this is a good place for the doctrine of double effect to come back into play. As a deontic principle, the DDE had problems, but as an aretaic assessment, it may be able to help explain why we want to divert the trolley; though it is the wrong thing to do. If we recall, the DDE states something like: all else held equal, there are stronger reasons against bringing about harm as a means to an end than there are against bringing about harm as a side effect. Used to assign blameworthiness, it would allow us to say that an agent would be due more blame for harms they intend to cause than they would be for harms that they bring about by accident.

Using the DDE in an aretaic way can help us explain why we find it so attractive, even though it is apparently flawed. One of the most attractive features of the DDE is that it takes an agent's intentions into account, and though it was demonstrated that reasons for performing a certain action should not have an effect on whether that action is right or wrong, there is reason to think that an agent's intention has something to do with his blameworthiness. Consider a variation of a case we have seen before:

> Transplant 3: A doctor has five patients, each dying from the failure of a single organ. One day, a healthy young person comes into the clinic for a simple check-up. As it turns out, this healthy young person is a match for all five of the doctor's dying patients. But, since one of the people dying

3 We also tend to think that the person who flips the switch has no one to blame but himself.

> from organ failure is the doctor's mortal enemy, he decides to let the single healthy person walk away unharmed.

In this case, the doctor does what we would intuitively believe to be the right thing, but he does it for the wrong reasons. Even though his action is right, we are averse to the choice, indicating that he is blameworthy in the death of the five.

So an agent's intention is probably a third component of blameworthiness, but how does this relate to the DDE, and the diversion cases? It cannot be as simple as saying that someone is less blameworthy for diverting the trolley because they intend to save the five. If this were the case, then that same person would not be blameworthy for pushing the fat man off the platform, so long as they intended to save the people tied to the tracks. So the DDE's usefulness as an aretaic principle cannot be directly tied to the result that an agent wants to achieve through his actions, or they would be able to justify cases like the Fat Man, or Transplant 2, on the basis that their intention is to save the five. The aretaic DDE must be able to separate cases of intentional harm from those of collateral damage, or else we run the risk of falling back into consequentialism.

The question is: what makes cases of collateral damage different? Or, what is it that makes our intuitions about blameworthiness in the Classic case different from those in a case like the one Judith Jarvis Thomson has proposed (1985):

> The Loop: You are walking along the trolley tracks, minding your own business, when you come across a fork. You notice that some way down, the tracks come back together, and just after that, there are five people tied down. Then, you notice one out of control trolley. If you do nothing, the trolley will turn left at the fork before rejoining the other track and killing the five. However, if you flip the switch the trolley will instead take the right hand track. There is a fat man tied to this track, and if the trolley hits him, it will not continue on, to kill the five.

The difference between the cases is not the intention of the agent who does (or does not) flip the switch; in both cases the goal is to save five people. The difference between the cases is the role that the damage plays. That is, whether or not the death of the single person is necessary for the agent to achieve their goal. In the Loop case, the five can only be saved if the trolley is diverted into the fat man; if he was not there, flipping the switch would have no effect on the fate of the five. The trolley would just loop back onto the main track, and kill them. In contrast, the death of the single person in the Classic case is purely

coincidental; their presence on the sidetrack does not necessarily figure into the agent's calculations. Perhaps our aretaic form of the DDE is better off without any strict appeal to intention, and ought to instead be worded like this: it is more blameworthy to require that someone be harmed in order for your goals to be achieved than it is to cause someone to be harmed as a side effect of your actions.

Of course, the DDE is always going to be subject to the problem of closeness, and the aretaic form of it is no exception. The problem of closeness brings to light the enquiry of when, would something actually be prohibited, or in this case made blameworthy, by the DDE. The principle, (as I have given it above) enquires what exactly harming someone in order to achieve your goals looks like. Earlier in this paper, the Poison Gas Hospital was given as an example of a case that is intuitively wrong, but would nonetheless, be permitted by the DDE. Opponents could bring up that case again, and point out that the DDE still permits it, though in this case that permission only has to do with the manufacture of the chemical and the consequent death of the single person not being worthy of blame.

My response is this: the critics are right, and as an aretaic principle, the DDE will say that the doctor's choice to save the five in the Poison Gas Hospital case is not blameworthy. But, even though we may not blame the doctor for his choice, that lack of blame does not make the action right. In contrast, if the manufacture of the chemical in question required the death of the one; for example, if it could only be made within that one's body, this would be an example of a blameworthy (and wrong action). Closeness is not such an issue for the aretaic version of the DDE, because this assessment is not trying to talk about right or wrong. The aretaic DDE is only meant to help explain our intuitions about cases involving collateral damage, and why sometimes we may want to do things even though they are wrong.

We arrive to the conclusion. We began with consequentialism, and saw that concentrating only on achieving the best outcome can lead to choices that seem intuitively wrong. We need forward-looking deontic principles to help us make choices. The first is easy to accept: the principle of beneficence requires that we act in ways that help others. But, we cannot only concentrate on doing good things; we must also try to avoid doing things that are wrong. So, we must balance beneficence with a principle of non-malfeasance that prohibits us from either causing, or deliberately failing, to prevent harm. Nonetheless, these principles fail to account for the intuitions that we have about cases of collateral damage like the Classic trolley, where we want to divert harm (even though someone else will be harmed by our doing so). This is where the doctrine of double effect was first introduced, but it ultimately failed as an

action guiding principle. It does, however, work as an aretaic one; our intuitions about collateral damage can be explained in terms of diverting not being blameworthy. Ultimately, the death of the ones on the sidetrack is not a necessary part of our plan. It is important to note that the action of diverting still violates the principles of non-malfeasance, it is still wrong; but, other factors mean that we would not blame a person for choosing to flip the switch.

BIBLIOGRAPHY

[1] Allowing, *The Philosophical Review* 98 no. 3. 287-312.

[2] Foot, Philippa. 1978. *Virtues and Vices*. Basil Blackwell: Oxford.

[3] Jarvis Thomson, Judith. 1985. 'The Trolley Problem', in *The Yale Law Journal* 94 no. 6. 1395-1415.

[4] Quinn, Warren. 1989. 'Actions, Intentions and Consequences: The Doctrine of Doing and Allowing', in *The Philosophical Review* 98 no. 3, 287-312.

[5] Rachels, J. 1975. 'Active and Passive Euthanasia', in *New England Journal of Medicine*. 292. 78-80.

[6] Unger, Peter. 1996. *Living High and Letting Die: Our Illusion of Innocence*. Oxford University Press: New York.