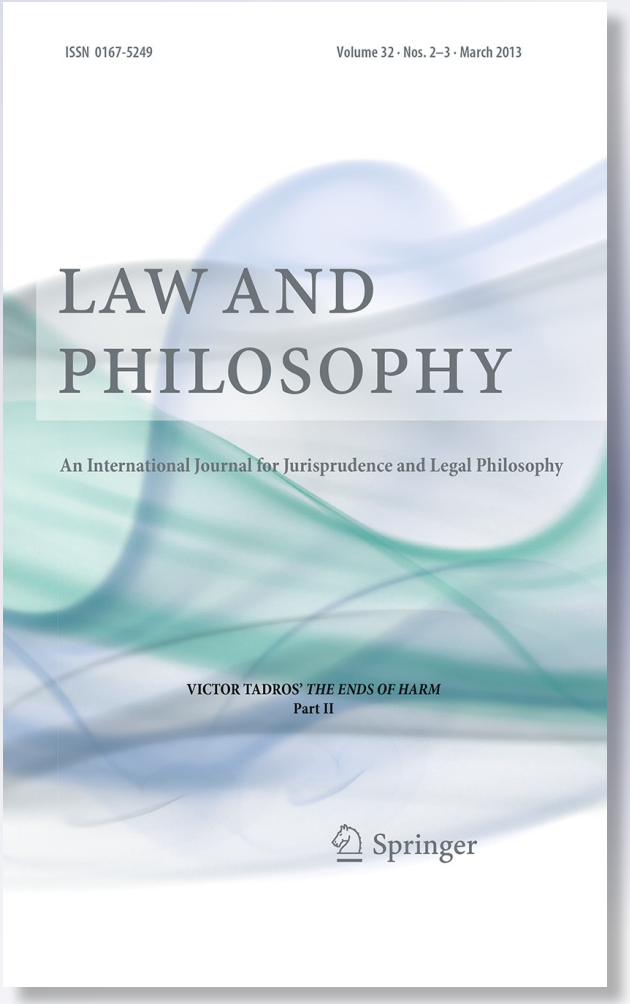
Wrongdoing Without Motives: Why Victor Tadros is Wrong About Wrongdoing and Motivation

Alec Walen

Law and Philosophy
An International Journal for
Jurisprudence and Legal Philosophy

ISSN 0167-5249
Volume 32
Combined 2-3

Law and Philos (2013) 32:217-240
DOI 10.1007/s10982-012-9154-1



Springer

Law and Philosophy (2013) 32: 217–240
DOI 10.1007/s10982-012-9154-1

ALEC WALEN

WRONGDOING WITHOUT MOTIVES: WHY VICTOR TADROS IS WRONG ABOUT WRONGDOING AND MOTIVATION

(Accepted 8 September 2012)

ABSTRACT. A central principle in Victor Tadros's book, The Ends of Harm, is the means principle (MP) which holds that it is, with limited exceptions, impermissible to use another as a means. Tadros defends a subjective, intention-focused interpretation of the MP, according to which to use another as a means is to form plans or intentions in which the other serves as a tool for advancing one's ends. My thesis here is that Tadros's defense of the subjective interpretation of the MP is unsuccessful. To make that case I argue for three claims. First, the subjective interpretation has implausibly harsh implications in certain cases, implying that certain people would be guilty of much more serious wrongs than they can plausibly be thought to have committed. Second, the cases that Tadros offers to argue that the subjective interpretation of the MP must be right are better interpreted as showing that it is impermissible to act on an illicit intention – one that would direct an agent under certain, foreseeable circumstances to perform impermissible acts – than that it is impermissible to act for an illicit reason. Third, while Tadros correctly rejects the objective, causal-role-focused interpretation of the MP – according to which to use another as a means is for the other to play the causal role of means to the good which might be offered to justify the act one performs – there is another way of defending the significance of causal roles, one that has implications that track those of the MP fairly closely. I argue elsewhere at length for this other principle, which I call the Restricting Claims Principle. Here I simply sketch the basic idea in a way sufficient to show that one can escape the dilemma that the MP faces without grabbing either the subjective or the objective horn, and without moving into a consequentialist world in which it is permissible to punish the innocent for the sake of the general welfare.

* I want to thank Victor Tadros for his insightful discussion of my arguments here, and Kim Ferzan for organizing the symposium and for encouraging me to focus on this particular feature of Tadros's book.

I. INTRODUCTION

The general idea of the *means principle* [MP] is that whilst it may be permissible to pursue the good where this will have, as one of its side effects, some lesser harm to others, it is not permissible to pursue the good where others will be used as a means to achieve that good.[1]

This is how Victor Tadros introduces the principle at the center of his recent thought-provoking book on punishment theory. The MP is central to the project of his book because Tadros wants to give a justification of punishment (a) that respects the idea that the innocent may not be used merely as a means of improving the welfare of others,[2] while also allowing (b) that the state may harm those who have committed crimes insofar as the state can thereby use them to deter would-be criminals (themselves or others) from committing crimes. The first prong establishes that the view Tadros is advocating is not a consequentialist account of punishment. The second establishes that it is also an instrumentalist justification of punishment.

A core problem for Tadros, then, is reconciling the use of criminals as a means of deterring other criminals with the MP. The general shape of his solution is to point out that the MP has qualifications or limits. The most commonly expressed qualification – common in general philosophical discussions – is that it is presumptively impermissible to treat another as a means only when one treats him *merely* as a means. What it means to do that is a topic of some dispute, but the canonical way of treating someone as a means without treating him *merely* as a means involves consent. If A offers B his uncoerced, sufficiently well-informed consent to using him as a means, then B may, in most circumstances, use A as a means. Normally such offers occur in exchange for something of value to A, but they can also be made as gifts.

The exchange or gift models of using another as a means but not merely as a means will not help reconcile Tadros's instrumentalist

[1] Victor Tadros, *The Ends of Harm* (New York: Oxford University Press, 2011), p. 114. From hereon, page citations from this book will be put in parentheses after the quoted material.

[2] Tadros qualifies this by saying that the MP should be treated as a sort of moral 'multiplier on the significance of the harm' inflicted on a victim who is used as a mere means for the welfare of others (p. 128). Thus Tadros thinks it is permissible to use one merely as a means, thereby causing his death, to save one million.

conception of punishment with the MP. For that, Tadros invokes a different qualification, less often noted, namely that it may be permissible to use someone as a means, even without his consent, if one is enforcing an obligation that he owes. This is relevant to the criminal law because, according to Tadros: 'Offenders incur duties as a result of their offending. These duties are plausibly enforceable… [And offenders] may be harmed as a means to compel them to carry their duties out' (p. 3). Thus Tadros labels his account of punishment a 'duty view' of punishment (p. 12).

I confess that the duty view of punishment seems problematic to me. I accept the idea that one who 'has wrongfully created a threat… can [permissibly] be harmed to prevent that threat from occurring' (p. 138). I have trouble, however, with the move from that proposition to the claim that an offender can be proportionately punished, after he has already caused harm, to prevent others from harming independent sets of victims. But for purposes of this symposium, I leave that issue to others.[3] What I want to focus on here is Tadros's interpretation of the MP.

Tadros comes to defend his interpretation of the MP in the context of addressing a dilemma. There are two plausible interpretations for what it could mean to use another as a means:

- a subjective, intention-focused interpretation, according to which to use another as a means is to form plans or intentions in which the other serves as a tool for advancing one's ends; and
- an objective, causal-role-focused interpretation, according to which to use another as a means is for the other to play the causal role of means to the good which might be offered to justify the action one performs.

Both, as Tadros notes, face potent objections. Nevertheless, Tadros opts for the subjective interpretation, and offers various arguments intended to show why that is the better choice.

It is important to address this dilemma because if Tadros were right to reject the objective interpretation, and if his defense of the subjective interpretation does not succeed – as I will argue is the case – then there would be good reason to think that the MP itself

[3] I offer my own explanation of why I think the duty view does not succeed in a review of *The Ends of Harm*, available at http://clcjbooks.rutgers.edu/books/ends_of_harm.html.

must be rejected. That in turn could have profound implications for criminal law theory, as it would suggest that the widely accepted hesitation to use the innocent as a means of promoting the welfare of others is the result of prejudice rather than defensible principle. And that in turn would open the door to argue for policies that many of us, who accept that the MP is either correct or onto something morally important, would view as monstrous.

My thesis here is that Tadros's defense of the subjective interpretation of the MP is unsuccessful, that the objective interpretation is facially implausible, but that there is another way of defending the significance of causal roles, one that has implications that track those of the MP fairly closely. Elsewhere I argue at length for this other principle, which I call the Restricting Claims Principle.[4] Here I will simply sketch the basic idea in a way sufficient to show that one can escape the dilemma that the MP faces without grabbing either horn, and without moving into a consequentialist world in which it is permissible to punish the innocent for the sake of the general welfare.

I proceed, then, as follows. First, I rehearse Tadros's account of the objections to a subjective interpretation of the MP. Second, I explore and reject Tadros's arguments for the subjective interpretation of the MP. Third, I offer a brief introduction to the Restricting Claims Principle, explaining why it serves as a superior deontic alternative to the MP.

II. OBJECTIONS TO THE SUBJECTIVE INTERPRETATION OF THE MP

The main reason to object to the subjective interpretation of the MP is that it directs the agent to focus on the wrong thing. As Tadros puts the point, 'we ought primarily to be focused outwards, on the people whose interests will be affected by our actions, rather than inwards, on our own motivations' (p. 146).[5] Tadros illustrates that point in the context of the notorious *Trolley Problem*, which is the challenge of explaining why it is permissible to kill one person,

[4] 'Transcending the Means Principle' (manuscript available at available at http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1372416).

[5] Tadros fails to give credit where credit is due for this argument. The seminal writings are: Jonathan Bennett, *Morality and Consequences, The Tanner Lectures on Human Values III* (Salt Lake City: University of Utah Press, 1981), p. 97; and J.J. Thomson, 'Self-Defense', *Philosophy and Public Affairs* 20 (1991): 283–310, p. 293.

without his consent, for the sake of five other people in the first but not the second of the following two cases.[6]

Trolley Switch: a bystander at a switch can throw the switch and thereby turn a trolley that is hurtling down a hill out of control away from five innocent people onto one innocent person on a side track (the 'side-track man'), thereby killing the side-track man and saving five.

Massive Man: a bystander at a switch can throw the switch and thereby cause a massive man to topple into the path of a trolley that is hurtling down a hill out of control, thereby killing him and saving five.

Most people believe that it is permissible for the bystander to turn the trolley, but not to topple the massive man in the trolley's path.[7] And Tadros adopts the initially plausible position that the MP explains why. The bystander would use the massive man as a means of saving the five, but she would not so use the side-track man. Rather, she would kill him as a side-effect of throwing the switch that turns the trolley.

Problems for the subjective interpretation of the MP can now be spelled out by considering some possible motivational states of the bystander at the trolley switch.

Case 1: suppose the bystander in *Trolley Switch* is not motivated to save the five, but is motivated by a desire to kill the side-track man. If she turns the trolley, does she act impermissibly? As Tadros notes, it is tempting to say instead that she 'has acted permissibly but for the wrong reasons' (p. 148).

Case 2: suppose the bystander is ambivalent and asks for advice about what to do. As Tadros says: 'Some people think it [would be] odd to say to

[6] The assertion that it is permissible to turn a trolley (or tram) from five onto one was first made by Philippa Foot in 'The Problem of Abortion and the Doctrine of the Double Effect', reprinted in P. Foot, *Virtues and Vices* (Berkeley: University of California Press, 1978), p. 23. But the second of the two cases and the label were introduced by J.J. Thomson in 'Killing, Letting Die, and the Trolley Problem', *The Monist* 59 (1976): 204–217. Tadros uses slightly different cases. His first case, which he calls 'Trolley Driver', involves a driver of the trolley, rather than a bystander. This case invites misunderstandings of what is at stake, as it is tempting to say that the driver faces a choice between killing and killing, whereas the bystander in *Massive Man* clearly has a choice between killing and letting die. Tadros himself switches to the bystander version on p. 156. I also change the label in the second case so that both cases describe the means that could be used to save five.

[7] For just one example of the vast empirical literature supporting this claim, see F. Cushman, L. Young and M.D. Hauser, 'The Role of Reasoning and Intuition in Moral Judgments: Testing Three Principles of Harm', *Psychological Science* 17 (2006): 1082–1089.

[her]: 'Well, that depends on what your motivations will be. If you act in order to save the five[,] you would be acting permissibly, but if you would be intending to kill the one, you would be acting wrongly'' (p. 149).[8]

Case 3: Suppose the bystander is badly motivated but skillful at throwing trolley switches. And suppose that I am well motivated but not so skillful; I might fail in my attempt to throw it. 'Wouldn't it be odd for me to prevent [her] from turning the trolley so that I, with my better motivations, could turn it in exactly the same way…?' (p. 149). As Tadros goes on to say, 'many people would think that I was overly concerned with motivations and insufficiently concerned with the saving of the five' (p. 149).

Another way of putting the point that these cases are meant to support – a way that Tadros does not mention – is that an 'intention is wrongful because the act intended is wrongful'.[9] Some other set of considerations must determine the more fundamental point, that the act itself is wrongful. This is obviously too strong to cover all cases; sometimes intentions do matter for the permissibility of actions.[10] But in cases like *Massive Man*, this seems to be the right thing to say.

III. TADROS'S ATTEMPT TO REHABILITATE THE SUBJECTIVE INTERPRETATION OF THE MP, AND RESPONSES THERETO

Tadros has a multi-pronged defense of the subjective interpretation of the MP. I will discuss two main prongs, and then wrap up by quickly responding to two ancillary arguments. But I should say up front that I think the first prong I discuss contains a significant mistake, which is, in truth, sufficient to show that Tadros's project fails. The rest of the discussion is meant simply to show that he has presented no other reason to think that we face a dilemma in which some reasons seem to show that the subjective interpretation must be wrong while others show that it must be right. In truth, nothing shows that it must be right, and there are decisive reasons to think it must be wrong.

[8] The best source for 'some people' is J.J. Thomson. See 'Physician-Assisted Suicide: Two Arguments', *Ethics* 109 (1999): 497–518, pp. 514–515. Thomson thinks it would not only be odd but 'an absurdity' to say what is said in the text. *Id*. p. 515.

[9] T.M. Scanlon, *Moral Dimensions* (Harvard University Press, 2008), p. 29.

[10] Among the examples Scanlon offers are what he calls 'expression and expectation' cases. *Moral Dimensions*, pp. 39–40.

A. Intending Wrong is Key to Wrongdoing

Tadros represents the idea of a permissible act as one that is 'an option for a person who is properly motivated' (p. 156). On that analysis, turning the trolley in *Trolley Switch* is permissible for the following reasons. First, one can be properly motivated and knowingly cause a harm as long as one does not intend to cause that harm, but rather acts only foreseeing that one will cause a lesser harm as a side-effect of creating a greater good. Second, one can turn the trolley only foreseeing, rather than intending, that one will cause a lesser harm. Therefore turning the trolley is an option for one who is properly motivated.

It follows on this analysis that what makes an act impermissible is that it could only (knowingly) be performed by a person who is not properly motivated. In the context of the MP, acts like toppling the massive man in front of the trolley are impermissible because they could not be performed without intending (unjustifiable) harm to another, either as an intrinsically worthwhile end, or as a means to an intrinsically worthwhile end.

This analysis of the difference between permissible and impermissible acts naturally raises the question of what to say about the bystander in *Trolley Switch* if she is motivated to harm the side-track man. Tadros's answer is that '*[her]* turning of the trolley has a wrong-making feature: it was done with the intention of killing the [side-track] man' (p. 156). It is therefore impermissible.

The problem with this theory of wrongdoing is that it wrongly implies that a bystander who turns the trolley from five onto the side-track man for an illicit reason acts impermissibly in just the same way as a bystander who topples the massive man in the path of the trolley. In fact, insofar as this theory of wrongdoing holds that the badness of an agent's reason for action makes her action impermissible, it seems to imply that the bystander who turns the trolley out of malice has performed a more egregiously impermissible act than the bystander who topples the massive man in front of the trolley. For at least the bystander toppling the massive man is, we assume, acting for a good end; the one who turns the trolley out of hatred for the side-track man has only evil as her end. Nevertheless, it seems morally implausible to say that the malicious bystander in *Trolley Switch* commits murder, which is exactly what we want to say of the bystander in *Massive Man*.

To be clear, it would be reasonable to say of a person who turns a trolley onto the side-track man without knowing that she is also thereby saving five that she must be aiming to murder the one, and that her act is as egregiously wrong if not worse than the toppling of the massive man in front of the trolley. But if the bystander in *Trolley Switch* knows that she is saving five, and knows that this would be sufficient justification for turning the trolley *if she were properly motivated*, and furthermore if she knows that she is permitted *not* to turn the trolley if she doesn't want to, then her choosing to turn it for an illicit reason is just exploiting a moral loophole. It is her achieving an end she wants – the death of the side-track man – by means of an act that she knows is at least in some sense permissible. Such loop-holing is morally odious; she should not be positively motivated to kill the side-track man. But to hold that her act is wrongful on a par with murder is simply implausible.

Only two replies to this objection are possible. First, Tadros could claim that there is something else that makes it much worse to topple the massive man in front of the trolley than to turn the trolley onto the side-track man for an illicit reason. Second, he can bite the bullet and defend this counter-intuitive view. I discuss both possibilities below.

Tadros suggested to me in personal communication that the fact that the loop-holing bystander's act in *Trolley Switch* is the *kind of act* that could be permissibly performed by an agent acting on good motives shows that it is not as egregious a wrong as the bystander's act in *Massive Man*, an act that can only be performed by an agent intending to harm another. Since the kind of act a loop-holing bystander performs is less bad, her performance of that impermissible act is a less egregious wrong.

The problem with such a move is that the 'kind of act' notion it presupposes cuts across the grain of the significance of intentions. The thing that unites illicitly motivated and well motivated bystanders in *Trolley Switch* as agents who perform the same kind of act is that they perform the same *physical* act, an act that leads to harm in a particular *causal* pattern, one that is distinct from the *causal* pattern in *Massive Man*. The subjective interpretation of the MP, however, suggests that we should stay focused on intentions, rather than causal roles. If we stay focused on intentions, we see that the

natural kind grouping is between the illicitly motivated bystander in *Trolley Switch* and the bystander in *Massive Man*. Both of them act on the intention to harm someone. Thus, if we want to stay within the framework of the subjective interpretation, we cannot appeal to the 'kind of act' notion that would seem to distinguish the loop-holing bystander in *Trolley Switch* from the bystander in *Massive Man*.

One might object that I am wrongly relying on the thought that the subjective interpretation must appeal to intentions and nothing but intentions to distinguish permissible from impermissible acts, and worse from less bad impermissible acts.[11] But that is not true. My point is only that when all other variables are held constant, as they are in the *Trolley Problem* cases, the explanatory work must either be done by the intentions of the agents or the causal roles of the patients. If the subjective interpretation of the MP has to help itself to something else, it seems to help itself to its competitor account, the causal interpretation, thereby undermining its own plausibility.

Second, Tadros could bite the bullet and assert that it is *just as bad, if not worse*, to turn the trolley for an illicit reason as to topple the massive man in front of the trolley. But this would be implausible indeed. Even the most extreme hate crime legislation still presupposes that the offender would have committed a crime even if she had not been acting with hatred of her victim. Acting on hate is meant to be only a sentence enhancer, not something that turns a legal act into a crime. Were Tadros to bite the bullet on this score, he would be taking the position, in the context of the *Trolley Problem*, that it is permissible to punish someone who performs an otherwise permissible act for an illicit reason as though she had committed murder. No one in either law or philosophy, as far as I know, has ever tried to advance such a position.

Tadros nonetheless offers some hint of wanting to bite the bullet in this way. He says that he believes that the bystander who turns the trolley for an illicit reason 'is liable to be punished' (p. 165). If he were willing to say that she is liable to be punished as much or more

[11] Tadros also suggested to me, following Warren Quinn, that not all instances of intentionally harming are equally bad. But even if this is true, it cannot help Tadros. Manipulative harming (Quinn called it opportunistic harming) is supposed to make terror bombing worse than the certain kinds of abortions (those involving craniotomies), which involve eliminative harming. But Quinn's distinction works on the assumption that the actor is aiming at a good. If in the craniotomy case the agent aimed directly at harm, that would be *even worse* than manipulative harming aiming ultimately at some good.

than the bystander who topples the massive man in front of the trolley, then he would indeed be biting the bullet.

As a matter of fact, Tadros does not want to go that far. He goes on to give 'good reason to refrain from condemning and punishing [her]', namely that doing so 'would discourage other bad [bystanders] from acting in a way that leaves the five to be saved' (p. 165).[12] This may seem to indicate that he does not want to bite the bullet. But it is not a clear indication of that. The position that there are important instrumental reasons not to punish loop-holing trolley turners is consistent with saying that they are liable, instrumental considerations notwithstanding, to being punished as murderers.

In addition, even if Tadros wants to defend his approach by saying that he ends up with a reasonable position – that a loop-holing trolley turner should *not* be punished for murder – he gets to this result in an implausible way. The better position is that loop-holing trolley turners are not culpable on a level approaching murder. They act on illicit reasons and deserve blame, but they cause no great wrong. The death they cause is objectively justifiable. The greatest wrong they cause is insult or offense. Such offense cannot be equated, in terms of the magnitude of wrong done, to the wrong of killing someone who has a right not to be killed.

One might object that people have a right not to be killed *for bad reasons*. Perhaps. But if they have no right not to be killed in exactly those circumstances, the wrong done is, again, the wrong of an offense, which is a categorically less serious wrong than the wrong of being killed in circumstances that do not justify killing at all. That is, at bottom, why it is morally implausible to bite the bullet and say that the loop-holing trolley turner is culpable for murder.

B. Duress, the Poisoned Pipe Case, and the Doctrine of Illicit Intentions

Tadros offers a set of cases meant to show that my last point must be wrong, that reasons for action certainly do distinguish between those who kill permissibly and those who murder. These cases concern the defense of duress and an interesting case of joint activity, the *Poisoned*

[12] I accept Tadros's argument that it is permissible to encourage another to act impermissibly when one is trying to achieve something permissible through the agency of another (pp. 160–162). I took a similar position in Alec Walen, 'Permissibly Encouraging the Impermissible', *Journal of Value Inquiry* 38 (2004): 341–354.

Pipe case. Together they successfully illustrate the fact that intentions sometimes matter to the permissibility of actions. But they do not illustrate that the MP should be given a subjective interpretation. That is, they do not illustrate that illicit reasons are a serious wrong-making feature of actions. Instead, they illustrate the need to adopt what I have elsewhere called the Doctrine of Illicit Intentions (DII).[13]

To explain, I start by giving Tadros's two cases, duress first. I then explain how the DII can handle these cases. I then offer two objections that Tadros was kind enough to share with me on the basis of an earlier draft of this article, and give my reply to his objections.

(i) *Tadros's two cases*: His point about duress is two-pronged. First, if the threatened harm to the agent is very great, and the harm the agent is asked to cause is fairly small, then an act that would normally be a wrong may be justified and permissible. His example is set in Nazi Germany. We are to suppose 'that the only way of averting a grave risk of having a family member being sent to the concentration camp was to help distribute Nazi propaganda' (p. 158). If the propaganda would make little difference to the Nazi cause, then this act seems morally permissible. Second, he points out that the act is permissible only if one were motivated by such a threat. It would be morally impermissible to distribute Nazi propaganda if one were doing so because one wanted to help promote the Nazi cause.

It might be tempting to say that it is not the intention of the agent distributing the propaganda, but the fact of the threat, that makes it permissible for one who has been given such a threat to distribute the propaganda. But Tadros has a clever response to that move. He imagines a gang of criminals who have each pledged to kill the family members of anyone in the group who tries to back out of group activities. He then imagines that the group has decided to rob a post office. We are to suppose that almost all members of the gang think that this is a good idea, but one member has decided that she does not want to engage in criminal activities. Still, given the threat to her relatives, she engages in the robbery. She should be able to make use of the defense of duress; the others, despite facing the same threat, should not.

[13] Alec Walen, 'The Doctrine of Illicit Intentions', *Philosophy and Public Affairs* 34 (2006): 39–67.

Tadros's second case involves two people, A and B, who independently put poison into the water pipe leading to a victim, V's, house – the *Poisoned Pipe* case. They each put in sufficient poison to kill V, and they are each aware of the other's actions. Moreover, 'A's poison alone or B's poison alone would lead to a very slow and painful death for V. Their poison together kills V swiftly' (p. 159). If they were both acting trying to kill V then they are both acting wrongly. But we can imagine that B acts with benevolent motives. She can't stop A or in any other way save V, but she can add her poison to the water to provide V a quicker, more painless death. It is at least plausible that the beneficent B is acting permissibly.[14] If so, then in this case, as in the duress cases, what is permissible depends on the intention of the agent.

The question is what to make of these cases. Do they show that acting for an illicit reason itself sometimes suffices to make an action egregiously wrong? Or is there some alternative reading of them? If they show that acting for an illicit reason is itself a serious wrong-making feature of action, then it is puzzling that the illicitly motivated bystander at the trolley switch should seem to be acting permissibly, or at least not committing a crime anywhere near as serious as murder. The DII provides another way of making sense of these cases, one that avoids this puzzle.

(ii) *The DII*: This is not the place to spell out the DII in detail. What I can do is spell out the basic idea and show how it applies to these cases. A core distinction the DII makes is between an illicit reason and an illicit intention. An illicit reason is a reason for action that flouts the limits of moral permissibility. The bad bystander in *Trolley Switch*, who aims to kill the side-track man, acts for an illicit reason. An illicit intention, on the other hand, disposes an agent, in ways that depend on the circumstances she finds herself in, to perform impermissible actions. The bad bystander who would only turn the trolley given that she knows that it is permissible for one in her position to do so does not have an illicit intention, even though she acts on an illicit reason. She would have an illicit intention if and only if her intention would direct her to kill the side-track man even when it is impermissible to do so.

[14] I am willing to accept, for the sake of argument, that this would be permissible. But there are good reasons to doubt it. It would not be permissible to poison someone, without her consent, who is dying a slow, painful death of cancer. It is hard to see why this situation should be different.

The DII holds that it is impermissible to act on an illicit intention, even though it is often permissible to act on an illicit reason.

It is important to note a few things about the DII. First, '[t]he role of intentions in determining whether actions are permissible is … secondary on this view'.[15] It presupposes 'an independent account of the kinds of actions that cannot permissibly be performed'.[16] What it adds is the thought that it is not only impermissible to perform impermissible actions, it is also impermissible to form and act on an intention to perform such actions.

Second, the DII takes into account that intentions often are not simple linear affairs with the shape: Do act A to achieve goal G. They are often complex affairs with a shape more like this: To achieve goal G, do A1 in condition C1; do A2 in C2; do A3 in C3, etc.; and if you have done A1, then do A1a in C1a, do A1b in C1b, etc.; and look out for conditions K1–Kn, in which case do nothing to pursue G. I call the range of actions that an intention might direct one to perform an intention's scope. The thought behind the DII is that an agent normally has no good reason to form an intention with a scope that includes impermissible actions.[17] Because an agent can have no good reason to include impermissible acts within the scope of her intentions, it is reasonable for morality to require her not to form intentions with impermissible acts in their scope.[18]

Third, it might seem that the DII directs one to focus inwardly on one's intentions just as much as the subjective interpretation of the MP does. But this is not true. In fact, the DII directs agents to focus only on what they will do. It involves their intentions only insofar as it directs them to commit themselves to framing intentions that exclude impermissible acts from their scope. It does not matter how they do that; it matters only *that* they do that. In contrast, the subjective interpretation of the MP does require agents to think about their reasons for action. It tells someone like the bad bystander at the trolley switch not to change what she would *do*, but to change

[15] 'Doctrine of Illicit Intentions', p. 39.

[16] *Id.*

[17] One possibility I overlooked when I published the 'Doctrine of Illicit Intentions' is that one might need to sincerely threaten to do something impermissible to get someone else not to do something impermissible, and this threat may be permissible. This is the only exception I am now aware of.

[18] It may be difficult to discern whether one's intention would direct one to perform impermissible actions. To accommodate that fact, the DII holds that one acts permissibly as long as one has not been reckless or negligent in failing to exclude those possibilities.

the reasons for which she would do it. In that way the two doctrines are completely different. Yet the DII still concerns intentions because it gives us a ground for condemning someone who performs an otherwise permissible act while acting on an illicit intention. It says not that the final act itself was impermissible, but that acting on an illicit intention, one that could well have directed her to perform an impermissible act, was itself an impermissible choice.

(iii) *Applying the DII to Tadros's two cases*: Now we can apply this to Tadros's two cases. Start with the case of the gang members, one of whom is robbing a post office under duress. The one who is acting under duress is acting on an intention whose scope seems to contain no impermissible acts. We can assume that the only reason she is robbing the post office is to protect her family, a condition under which that action is permissible. If freed from such coercive threats, her intention would presumably not direct her to engage in post office robbery or any other impermissible actions. By contrast, her fellow gang members are ready and willing to engage in post office robbery. Even if the mutually enforcing threat were to drop away, they would still rob the post office because they think the money they get is justification enough for them. Thus they act on illicit intentions; she does not.

The same thing is true of beneficent B in *Poisoned Pipe*. Assuming that her act is permissible, it is permissible only in this odd circumstance in which by adding poison to the pipe, she helps V suffer a less bad death. Beneficent B would add poison to the pipe only in such a highly restricted circumstance, so her intention is licit. By contrast, her maleficent counterpart would perform the act of poisoning whether it would help V or not. She then clearly acts on an illicit intention, and that is why she and not her beneficent counterpart should be prosecuted.

(iv) *Tadros's objections and replies*[19]: Tadros raised two objections to my use of the DII to handle his cases; we can explore them by focusing on my treatment of *Poisoned Pipe*. First, he said that I can't distinguish someone who would order coffee willing to shoot the barista if she doesn't sell him the coffee, but ready to pay normally if she does, from the bad actor in *Poisoned Pipe* who kills for the money:

[19] This and the next subsection are taken, with some changes, from a blog entry I posted on Pea Soup called 'Intentions and Permissibility,' available at http://peasoup.typepad.com/peasoup/2012/04/intentions-and-permissibility.html.

the acts each actually performs are justified, but both are ready to perform acts that aren't justified. Yet surely the bad actor in *Poisoned Pipe* is a murderer, while the coffee customer has committed no crime. Second, he says that I'm just getting the wrong in *Poisoned Pipe* wrong: it's killing, not acting on an illicit intention.

Here I must pause to say how grateful I am to Tadros for raising these objections, as they pushed me to develop the DII in ways I had not seen the need to do before. That said, I now have a view about how the view should be developed. To explain it, we need to start by distinguishing different states of awareness in the bad actor in *Poisoned Pipe*. First, assume the poisoner is not aware of the justification for the act – assuming again that the justification is successful. Assume, that is, that she is adding the poison to kill and get the money. I say that if she kills and was unaware of the justification for doing so, she is guilty of the equivalent of attempted murder. Why attempted murder and not murder? Because no death was caused that could not justifiably be caused. Thus murder does not fit. But from the agent's point of view, she was trying to cause an unjustified death. Thus she should be held culpable for the equivalent of attempted murder.[20]

Now what if she is aware of the justification but it was just a fortuity of the case; she would have killed anyway? This version of the case comes closer to the coffee case, as both agents know that what they do is permissible, yet both stand ready to do something impermissible. Nonetheless, I think we can still describe a significant difference between the levels of culpability of the two agents. This difference can be traced to the nature of the reason acted on. In *Poisoned Pipe* the reason is a murderous one: kill V to get some money. The agent got lucky, finding that in this case it turns out that the act is permissible. But this leaves her with a level of culpability equivalent to attempted manslaughter. Why manslaughter and not murder? Because at the time of her action, she realized that the act was justified. If acting without such knowledge should make her culpable for attempted murder, then she should be less culpable when she has that knowledge. But why is she culpable at all? Because forming that intention was reckless. At the time she formed it, she

[20] This is what Paul Robinson calls the 'deeds theory', as opposed to the 'reasons theory', of unknowingly justified acts. *See* 'The Bomb Thief and the Theory of Justification Defenses', *Criminal Law Forum* 8 (1998): 387–409. I agree wholeheartedly with Robinson's endorsement of the deeds theory.

was intending to murder. But for good luck, she would have killed impermissibly.[21]

In the coffee case, by contrast, the main reason to act is morally neutral: to get coffee. Acting on the illicit intention is still morally impermissible. That is, I think he can reasonably be required *not* to intend to kill, even conditionally, even if the condition is very unlikely to occur. But given that the odds that the barista will not sell him coffee are indeed low, and he does not expect to be in a situation where his intention would direct him to kill her, his culpability for adopting that intention seems sufficiently low that the law should be reluctant to criminalize it. I'd say that the law should criminalize acting on an illicit intention, grounded on a morally neutral or good reason, only if the person was expecting that he would likely perform an impermissible act or was actively taking steps to be ready to act impermissibly. Thus I have an account of the serious crime committed by the bad agent in *Poisoned Pipe*, and a response to the claim that I can't distinguish that case from the coffee case.[22]

Tadros can still object that my view is quite peculiar. On my view, those who kill while acting on an illicit intention in cases like *Poisoned Pipe* are guilty of nothing worse than *attempted* murder, or perhaps even attempted *manslaughter*. This just shows, he might say, that I'm getting the cases wrong. Again, I treat the wrong as one of acting on an illicit intention, but the wrong is *killing*.

Here I must acknowledge that my view is in this way counterintuitive. The majority of criminal law commentators seem to think that cases of unknown justification should be treated as if they were murder.[23] But the fact that my view is in the minority – and novel insofar as it extends to treating cases in which the agent stumbles

[21] Such a reduction in culpability might not be available in cases like the gang case, where people set it up that they have the reason to do the crime—much as one cannot escape mens rea requirements by getting drunk so that one doesn't know what one is doing. See *Model Penal Code* § 2.08(2) (1985).

[22] An editor at *Law and Philosophy* presented this hypothetical: 'Say that I go to buy a new expensive and sharp kitchen knife'. As I am about to purchase it, I think: 'this is good to buy because (1) I am having a dinner party tomorrow and I need it for the special salad I am making and (2) my husband is driving me nuts and I may decide to kill him with it'. Is this a criminal intention? It depends how we flesh out the case. If she buys the knife thinking: 'if he does X one more time, I'll kill him with this knife', and if it is likely that he will do X, then it could count as an attempted murder. But if her thought is literally 'I may decide to kill him with it', then though she has considered forming an illicit intention, she has not yet formed one, and therefore has formed no criminal intention. As I wrote before, 'an agent does nothing impermissible unless and until [she] forms or acts on an intention that could direct [her], without further modification, to perform independently impermissible acts'. 'Doctrine of illicit Intentions', p. 58.

[23] *See* Paul Robinson, 'The Bomb Thief and the Theory of Justification Defenses,' p. 392, nn. 12–15.

into an instance of known justification as a case of attempted manslaughter – does not show that it is wrong. To be clear, in saying that the killing in *Poisoned Pipe* is itself justifiable, and that it is the bad agent's acting on the illicit intention that is criminal, I am not suggesting that she is guilty of a thought crime, disconnected from her actions. I am saying that she was *acting* on an intention knowing that it could well direct her to murder. This is the essence of the culpability in attempted murder or recklessness, both of which are surely very serious crimes.

Finally, it is worth pointing out that Tadros's position is not that different. He too thinks the act is, in some sense, permissible, and that it is made impermissible because of the reason on which the person acts. Acting on an illicit reason, on his view, is the wrong-making feature of the action. That is not so different from the DII. Only, for reasons spelled out above, the subjective interpretation of the MP, with its focus on reasons for action, runs into serious problems, problems that the DII avoids. Thus even if the DII has this counter-intuitive implication, it is still the better of the two views.

C. The Significance of Attitudes

The last of Tadros's arguments for the subjective interpretation of the MP are actually a pair of arguments based on the thought that the MP must be concerned with the attitudes of agents. Both arguments fail because they fail to take into account the way that rights can work. Addressing these last arguments will serve as a segue to my positive account of how to make sense of the significance of causal roles.

The first of these arguments is that the MP reflects the moral status of people, and what it demands is that this status be *respected*. '[T]hat suggests [because respect is an attitude] that the attitudes of the person doing the harming will be important in explicating the [MP] (p. 153)'. My reply to this argument is that respect is to be shown to beings with the status of persons by respecting their rights. And as I will argue in the next section, rights give significance to causal role.

The very last of Tadros's arguments is based on a pair of cases that are designed to be analogs to the *Trolley Problem* cases, but that

involve asteroids, rather than persons, either diverting a trolley or toppling a massive man in front of the trolley. Tadros then says, '[i]f it were true that there were motivation-free facts that our motivations ought to track, we should expect our judgment about [these cases] to be similar to the judgment that we have about [*Trolley Switch* and *Massive Man*]' (p. 154). In truth, however, we don't react to the cases the same. When it's an asteroid acting, we think it equally sad that one had to die, and equally lucky that five get to live.

My reply to this argument is simple: asteroids don't violate rights. We should be committed to respecting rights. That does not mean that we have to value preventing rights violations more than we value preventing equivalent harms. It means only that we each have to accept that we are obliged not to violate rights. Thus each of us must not topple a massive man in front of a trolley, not even to save five. Moreover, each of us would have an obligation to save a massive man, if we could do so with minimal risk to ourselves, whether he had been toppled in front of a trolley by an asteroid or a person, even if saving him would allow five below to die. That is because he has a right to be saved as long as saving him would cost the agent relatively little; the competing claims of the others are to have him left as a resource for them, and they have no right to have him left for them in that way. Having this commitment not to violate rights, however, in no way implies that it is somehow *worse* if a harm befalls a person because he was wronged in a way that *causes* a good than if a harm befalls him as a *side-effect* of some good. The significance of causal roles is not to be found in a consequentialist analysis; it is rather deeply embedded in the structure of rights. It is to back up that claim that I turn, now, to my last section, explaining the Restricting Claims Principle.

IV. DEFENDING THE MORAL SIGNIFICANCE OF CAUSAL ROLE THROUGH THE RESTRICTING CLAIMS PRINCIPLE

We have seen that there are decisive reasons to reject the subjective interpretation of the MP, but Tadros is right to claim that the objective interpretation is equally implausible. The reason is simple: causal role has no obvious moral significance. Tim Scanlon put the point even more starkly: 'being a means in this sense – being causally

[24] Scanlon, *Moral Dimensions*, p. 118.

necessary – has no intrinsic moral significance'.[24] If we are, then, to avoid simply rejecting the kind of deontology that MP represents, we need some way of understanding why causal roles, pace Scanlon and Tadros, can be morally significant. I sketch here a way to do that, in the context of a theory of rights.

Before proceeding I want to note a general dispute about the value of rights as a conceptual framework on which Tadros and I differ. Tadros thinks that talk of rights is just a convenience, that can and often should be dispensed with.[25] As he says in one particular context:

> [B]y focusing on rights we mask rather than illuminate the deeper moral issues that should inform our decisions whether a person who attacks another retains their right to defend themselves against any defensive force used by the other person or not. … We would do better to move straight to the underlying considerations that determine what is permissible to do rather than using the language of rights in our explanation. (p. 201)

I accept that there are certain topics for which this is true, and I accept, in ways that I make concrete below, that rights talk must be based on deeper principles. But I disagree with the claim – only implicit in Tadros – that it is always possible to do away with the framework of rights. Rights do a specific kind of work directing different kinds of moral considerations into the right sort of balance – a balance that is not the same as the consequentialist benevolent observer balance, but that instead finds its fulcrum on the person of particular agents who have to make particular choices at particular points in time. That structure is what rights contribute, and it is easily lost by 'mov[ing] straight to the underlying considerations' (p. 201).

Turning now to the substance of the Restricting Claims Principle (RCP), its function is based on distinguishing restricting and non-restricting claims. It holds that the former are substantially weaker than the latter. The explanation for this difference rests on the fact that restricting claims have the potential, in a way that non-restricting claims do not, to restrict what an agent can do for others. As a result, restricting claims function, in a way that non-restricting claims do not, to provide normative pressure to make others worse off than they would be if the claimant were not present. In that sense

[25] In this regard he follows Joseph Raz's lead, though he does not say so explicitly.

they impose something like negative externalities on others. The RCP holds that restricting claims must therefore be weaker than otherwise identical non-restricting claims, to prevent those 'negative externalities' from being too large.

This maps fairly well onto the MP because those who would be used as a causal means of achieving a good have non-restricting claims not to be so used, while those who would be harmed as a side-effect of achieving a good have restricting claims against being so harmed. It follows that people have much stronger claims not to be harmed when they would serve as a means of achieving some good that might be offered to justify the act in question than when they would be harmed as a side-effect of pursuing such a good.[26]

This point can be illustrating by using the cases in the *Trolley Problem*. The massive man's claim is non-restricting because respecting his right not to be used as a means without his consent would not restrict the bystander from doing what she could otherwise, in his absence, do to save the five. Either there is something else she could permissibly do, such as turn the trolley onto another track, in which case respecting his claim not to be killed would not bar her from saving the five, or there is nothing else she could permissibly do, in which case respecting his claim as a right would not restrict her relative to the baseline of his not being present. In contrast, the side-track man's claim is restricting because, if the bystander had to respect it as a right, she would not be permitted to do what she could otherwise, in his absence, permissibly do for the five, namely turn the trolley onto the other track.

It is important to be clear that *all* claims on an agent have to be assessed in terms of whether they are restricting or not. Thus the claims of the five in both cases must also be assessed, and it turns out that they are restricting. Their being restricting follows from the fact that if the bystander had to respect their claims to be saved as rights, that would restrict her from doing what she could otherwise, if they were not present, do for the one, namely *not* turn a trolley in his direction in *Trolley Switch* or *not* topple him in the trolley's path in *Massive Man*. True, the side-track man's claim, if respected as a right,

[26] As I argue in 'Transcending the Means Principle', the relevant notion of causing or allowing harm, in the final analysis, must be framed against a baseline based on property notions that is itself thoroughly normative. Nonetheless, that fact need not concern us in this short treatment, as it is still a causal idea that sits at the core of the RCP.

would limit what she could do *for* others, while the claims of the five would limit what she could *not* do *to* others. But they are all restricting claims in the sense that they all have the potential, if treated as rights, to affect the welfare of others, compared to a background in which those with the competing claims are not present.

It is also important to be clear about how potential justifications operate to establish, in the place of intentions, the relevant causal relationships, and how rights operate with regard to potential justifications. A potential victim may have a right that would block one particular justification for an act, but not another. He has a right that an agent perform or not perform an act only if no justification for her doing the opposite succeeds. To illustrate the point, consider two possible justifications for turning the trolley onto the side-track man: one is to save the five, and the other is to kill the side-track man (say to revenge a personal slight). The side-track man has a right not to be killed for the second reason, but no right not to be killed for the first. The fact that he has no right not to be killed for the first reason means that the act of turning the trolley onto him is justifiable. Having separated the possible justifications from the corresponding intentions, we can also ask the separate question whether acting on the intention that corresponds to the invalid justification makes an act impermissible. If my arguments above were correct, the answer is that it generally does not (unless the intention violates the DII); it merely makes one subject to criticism for acting on a bad reason for action.[27]

With this much in view, we can now highlight the most fundamental difference between the RCP and the MP. The MP describes a moral fact about a dyadic relationship between an agent and a patient. It holds that there is a certain kind of action, using another merely as a means, that is morally extremely problematic. The RCP broadens the basic frame of relevant considerations to a global balance of claims on an agent. What matters for the permissibility of an agent's acts is the balance of the competing patient-claims on her,

[27] For a contrary view, see Matthew Hanser, 'Permissibility and Practical Inference', *Ethics* 115 (2005): 443–470. I respond to that view in 'The Doctrine of Illicit Intention', pp. 44–46.

[28] I discuss the significance of agent-claims in Alec Walen and David Wasserman, 'Agents, Impartiality, and the Priority of Claims over Duties: Diagnosing Why Thomson Still Gets the Trolley Problem Wrong by Appeal to the "Mechanics of Claims"', *Journal of Moral Philosophy* (forthcoming).

and her own agent-claims (I leave agent-claims to the side here).[28] There is still a dyadic relationship between an agent and every person who has a right against her. But those rights have to be understood as arising in the context of a larger balance of competing claims.

Having described the basic idea of the RCP, illustrated it, and distinguished it from the MP, I will now say a little bit more about the justification for it. I said above that one with a restricting claim imposes something like negative externalities on others. What I mean is that there is a meaningful patient (one affected by an agent) analog to the idea of a negative externality. A negative externality is a cost that an agent imposes on others by her actions. It is commonplace to say that agents should not be allowed to impose unjustified costs on others by their actions. The RCP is based on the idea that patients, while they do not *do* anything, can nonetheless also impose costs on others by being present with claims that have the potential to make others worse off than if they were not present – costs that show up in the balance of claims on an agent. Those costs, just like those of classical externalities, should not be allowed to be excessive.

Unless we treat restricting claims, whether negative claims not to be harmed or positive claims for aid, as more or less on a par with those of others with similar welfare interests – assuming there are no property rights in play – then one side will impose excessive costs on the other. For example, if we allow the side-track man's claim to count as a right that must be respected, his presence will cause the five on the other track to go from people who would have had a right to be saved to people who must be allowed to die. There is no obvious reason why his restricting claim not to be killed should win over their restricting claims to be saved. That is, there is no obvious reason why their claims could not just as well cause him to go from a man who must not be killed to one who must be killed. Of course, his claim is negative and theirs are positive. But they are alike in being restricting, and for reasons that I will spell out immediately below, that is a more important moral fact. Before explaining why that is so, however, it is worthwhile pointing out how different the

[28] I discuss the significance of agent-claims in Alec Walen and David Wasserman, 'Agents, Impartiality, and the Priority of Claims over Duties: Diagnosing Why Thomson Still Gets the Trolley Problem Wrong by Appeal to the "Mechanics of Claims"', *Journal of Moral Philosophy* (forthcoming).

situation is for the massive man. His claim can be respected as a right without making the others any worse off than if he were not present. Thus he and the five are not in the same sort of symmetrical position as the side-track man and the five. If the claims of the five could justify a bystander killing him, then their claims to be saved would make him worse off than if they were not present, despite the fact that his claim not to be harmed does not push to make them worse off than if he were not present.

Taking the justification one step deeper, the ultimate moral principle in play is that each person must be respected as one who has his own life to lead. That means that he cannot be asked to make great sacrifices for others as an agent, unless he has done something to the others, or stands in some special relationship with them, and thereby has acquired a duty to help them. And it likewise means that he cannot be asked to suffer a great cost on their behalf as a patient, again, unless he has done something to owe them what is extracted from him, or has claims that press to make them worse off than if he were not there. If he owes them a debt, then the debt (within the proper due process limits) may be extracted from him. And if his claims are restricting, then they have to be put in the balance with competing claims so that they register as roughly the same, modulo the welfare interests at stake and whatever difference the distinction between positive and negative claims makes. But if his claims are not restricting, and he owes them no debt, then he must be left more or less as unmolested as an analogous agent is unencumbered by the duty to sacrifice herself for strangers. Both the agent and the patient perspective are essential to having one's own life to lead.[29]

There are many details that this sketch does not cover. These include problematic cases of various sorts. I discuss them elsewhere at some length.[30] For present purposes, all I want to do is suggest that the RCP provides a meaningful alternative to the MP, one that makes it clear why a patient's causal role should matter morally. It matters because it determines the kind of claim a patient has in the balance of claims on an agent.

[29] Tadros, to his credit, is a rare author who gets this symmetry between the agent and the patient point of view (pp. 129–130). In 'Transcending the Means Principle' I argue that the symmetry is not complete, but it is nonetheless an important symmetry to observe.

[30] Again, see 'Transcending the Means Principle'.

In sum, those who would serve as a means of achieving a good that might justify an action that harms them are also people whose claim not to be harmed does not, if respected as a right, make others worse off than if they were not there. This means that they should be left unharmed unless the harm to them is more or less on a par with the kind of burden we could expect an analogous agent to assume for the sake of these same others. But if patients occupy the causal role of people who would be harmed as a side-effect of achieving a good that might justify an action that harms them, then their claims not to be harmed operate in a different way. They would, if respected as rights, make others worse off than if they were not there. Therefore their claims must not be regarded like those with non-restricting claims. Their interests are intertwined with those of the other patients, and their claims should register fundamentally on a par with those with whom they compete.

V. CONCLUSION

I have argued that Tadros is right to worry about the MP facing a dilemma. But he is wrong to think that the solution lies with the subjective interpretation of the MP. It lies, instead, with the RCP, which explains, inside a rights framework, why the causal role of being a means to an end that might justify a harm is morally significant. It is unclear if this mistake plays any role in distorting the rest of Tadros's theory of criminal punishment. I suspect it does not. Regardless, it is important to get a plausible account of what the MP is supposed to capture, because without it the door to consequentialist abuses of the individual is thrown wide open.

Rutgers University, New Brunswick,
NJ, USA
E-mail: awalen@warpmail.net