**The Permissibility of Deterrence**

Over the modern philosophy of punishment there looms the ‘Formula of Humanity’ in Kant’s Groundwork. This formula asserts, in part, that we may not use people simply as a means. This assertion has suggested to many philosophers, probably including Kant himself, that punishing people in order to achieve deterrence is morally objectionable in principle. What is called ‘general deterrence’ seems to run afoul of the Kantian principle: punishing a person to achieve it would involve imposing burdens on her in order to induce fear in other people that they will undergo such burdens if they break the law. This seems to be an example of using a person simply as a means to achieve a reduction in the incidence of crime.

To claim that deterrence is objectionable in principle is problematic, partly because deterrence has seemed to many people to be an obviously legitimate goal of punishment. The assertion also seems to rule out any use of consequentialist ideas in the design of institutions of punishment. Consequentialism and utilitarianism are often equated with the claim that the only legitimate goal of punishment is deterrence. This is false, but these theories certainly do support the claim that deterrence is an important and legitimate goal of punishment.[[1]](#endnote-1) And if the objection is correct it would also hold against any theory that claims that deterrence is even one of the legitimate goals of punishment.

In this paper I explore the degree to which the most plausible versions of a Kantian approach to punishment differ from plausible versions of a consequentialist approach with regard to the permissibility of deterrence. I begin by examining the Formula of Humanity. Perhaps surprisingly, I show that the most plausible statement of this principle does not even mention the idea of treating people merely as a means. The other crucial claim in that principle—that we must treat people as ends—is in fact the operative idea. The best interpretation of the Formula of Humanity makes Kantianism a basically ‘objective’ moral theory, in a sense to be explained, as is consequentialism. After defending my interpretation of the formula I go on to consider a recent theory of punishment developed by Victor Tadros that appeals to a principle paralleling the Formula of Humanity, the Means Principle. I argue that the Means Principle is not a plausible moral principle by using arguments that derive from my interpretation of the Formula of Humanity. The Means Principle is therefore unsuited to be the way to frame a discussion of whether deterrent punishments are permissible. In Section IV I discuss a version Kantian moral theory that can be seen as based on a proper interpretation of the Formula of Humanity. This version, drawing on the work of John Rawls, permits the pursuit of deterrence. This suggests that Kantianism and consequentialism are not divided on the issue of the permissibility of deterrence. In the final section I note one reason for thinking that further features of the consequentialist approach to deterrence make it superior to the Kantian-Rawlsian.

I

The Objection and Its History

The idea that punishment may legitimately be used to deter potential criminals from committing crimes is accepted by moral common sense, I believe. It is true that moral common sense accepts that there are some limits on the pursuit of deterrence. For one thing, at some point an increase in the severity of punishment, even if it deterred more potential criminals, would be rejected as morally unacceptable. So, for example, most people would object to imposing 30 years of prison on an offender who was convicted of stealing one laptop, even if it could be shown that doing this resulted in fewer incidents of this crime. However, I will mainly be examining the claim that the pursuit of deterrence is objectionable in principle.

There is a second argument related to deterrence that I will also not be examining. This is the well-known objection to consequentialism that asserts that this theory must grant that it is sometimes morally permissible to punish someone known to be innocent.[[2]](#endnote-2) However, the ‘in principle’ objection to deterrence applies even to the punishment of the guilty.

The objection in principle might well be thought to have originated with Kant. There is some plausibility in this attribution. And there is, to be sure, some intuitive plausibility in the objection itself, whoever deserves the credit for it. The attribution of the objection to Kant obviously stems from the way that he sometimes formulates the supreme principle of morality, the Categorical Imperative. Let us begin with the following statement of the relevant version, the Formula of Humanity.

FH: Act in such a way that you always treat humanity, whether in your own person or in the person of any other, never simply as a means, but always at the same time as an end.[[3]](#endnote-3)

This asserts, in part, that it is morally wrong to treat people simply as a means. The objection would then amount to the claim that in punishing a person to achieve deterrence we are doing precisely that. Consider ‘general deterrence’, that is, punishing an offender who has committed a crime in order to frighten other people off from doing so. If this is our aim in punishing the offender—the objection goes—then we are inflicting pain or a loss on her simply as a means of influencing the behavior of other people. She is being used like a scarecrow, a mere thing, and this, it could plausibly be said, is morally wrong.

Did Kant himself actually make this objection? Scholars have debated the answer to that question. The most explicit discussion of punishment by Kant in a published work is a brief passage in the first part of The Metaphysics of Morals. He says this:

Punishment…can never be inflicted merely as a means to promote some other good for the criminal himself or for civil society….For a man can never be treated merely as a means for the purposes of another or be put among the objects of rights to things: His innate personality protects him from this…[[4]](#endnote-4)

This certainly seems to be stating that we may not treat a person like a scarecrow, simply to further the ends of others.[[5]](#endnote-5) But, as A. C. Ewing noted, it is also stating—rather chillingly, I think— that we may not inflict punishment on an offender in order to promote his own good. Ewing’s point brings out that whatever argument Kant is making here it seems not to be directed only at general deterrence or, indeed, only at forms of deterrence. (Ewing takes Kant to be objecting to ‘reform’ theories as well.) Furthermore, there are scholars who believe that Kant elsewhere does grant at least some role for deterrence in his larger political and legal theory.[[6]](#endnote-6)

The argument that Kant seems to make has appeared often in the literature on punishment since the 1970’s, which gives it a markedly retributivist orientation. A number of important contemporary writers, such as Antony Duff, Andrew von Hirsch, and (at one time) Jean Hampton endorse the argument, or something closely-related.[[7]](#endnote-7) What is more, it seems to be implicitly accepted by Robert Nozick and Michael Moore, among others, who try to develop entirely retributive theories of punishment where deterrence has no role.[[8]](#endnote-8)

Finally, there is something like the Kantian argument in the most substantial recent work in the philosophy of punishment, Victor Tadros’ The Ends of Harm. In this work Tadros develops a theory of punishment, the ‘duty view’, that he claims is neither retributivist nor consequentialist. Tadros asserts that a certain amount of general deterrence is morally permissible, so he does not object to it in principle. But it is not permissible on consequentialist grounds. In other words, he accepts that punishment may be designed to achieve the goal of deterrence, but this permission derives from a non-consequentialist moral theory.

Tadros accepts what he calls the Means Principle, and he uses this to object to consequentialism. This principle can best be formulated as follows.

MP: If an action harms a person [simply] as a means of producing good for other persons then it is not permissible.[[9]](#endnote-9)

MP has some similarities to the Formula of Humanity (FH), and Tadros reinforces our sense of the connection by arguing that MP can be supported by appealing to the Kantian idea that persons have a special moral status. This status, he says, demands that we not be regarded as available for the use of others. Echoing Kant, Tadros says that we are not so available because “as humans, we are capable of determining what ends to set for ourselves”.[[10]](#endnote-10)

Tadros does not believe that MP is absolute; he thinks that there are cases where a person may permissibly be harmed in order to benefit others. And he argues that in some cases deterrent punishments can be shown to be permissible because they fall under such an exception to MP. But MP puts the practice of punishment under a serious cloud of suspicion, and it is said to be based on Kantian ideas about the permissible treatment of rational agents. I will consider whether Tadros can use MP to rule out the pursuit of general deterrence in the design of a system of punishment—unless we can find an exception to it that a non-consequentialist moral theory can allow for.

II

Treating Simply as a Means and Treating as an End

In this section I consider how to interpret the Formula of Humanity. I investigated this question elsewhere.[[11]](#endnote-11) The conclusion I reached, drawing on the work of Kant scholars, was this: FH is best understood as requiring us to always treat rational agents as ‘ends’, in Kant’s technical sense; the prohibition on treating people simply as means does not establish which actions are right or wrong. This entails that whenever it is morally wrong to treat someone simply as a means the explanation for this fact is that doing so fails to treat the person as an end.[[12]](#endnote-12) FH establishes a largely objective moral theory, in a sense I will explain—as does consequentialism.

We can begin by recalling that Kant famously distinguished between acting ‘in conformity with duty’ and acting ‘from duty’. And he went on to assert that a merchant, who returns the correct change to someone from self-interest nonetheless acts rightly.[[13]](#endnote-13) In everyday language we make a similar distinction when we say that someone did the ‘right thing for the wrong reason’.

These kinds of claims suggest that what we can call the deontic status of an action is largely ‘objective’ in a technical sense. The three deontic statuses that an action may have are obligatory, wrong, and permissible. The deontic status of an action is objective in the relevant sense if it does not vary with the subjective features of the agent as she acts, such as her motive or beliefs.

The motive of an action, I have argued, is an agent’s ultimate desire in acting; that is, the desire which establishes an agent’s ultimate goal in acting.[[14]](#endnote-14) It is to be distinguished from derivative desires. For example, suppose that S wants to relieve an itch and she believes that her scratching it will relieve her discomfort, so that she comes to desire to scratch it. If she scratches the itch in order to relieve her discomfort, her motive for doing this is her desire to relieve her discomfort. Her desire to scratch the itch is derivative; it is not her motive.

As I stated, we generally think that the motive of an action does not affect whether it is right, wrong or merely permissible. If I tell a lie, for example, that is generally wrong, and it is wrong whatever my motive is for telling it. And, on the other hand, as Mill noted, if I rescue someone from drowning that action is right whatever my motive is for doing so.[[15]](#endnote-15) There may be rare cases where the motive of an action is relevant deontically—I believe that there are—but otherwise motives are irrelevant deontically.[[16]](#endnote-16) Most familiar moral theories accept that deontic status is largely, or even entirely, objective. Mill implies that deontic status is entirely objective in utilitarianism when he says that we must distinguish “a standard of morals” or a “rule of action” from the motive of it.[[17]](#endnote-17)

The relevance of these claims to our investigation can be explained as follows. We are trying to understand what Kant thinks the fundamental principle of morality is. Such a principle will establish the deontic status of all actions. It should thus yield a largely objective conception of deontic status, as Kant himself seemed to recognize. In other words, the principle should yield deontic judgments that make motives irrelevant in most cases.

Now the means that a person employs in achieving her goals do not constitute her motives. Her motives set her goals or ends in acting, and a desire to employ a given means is derivative, not ultimate. But it remains the case that a largely objective conception of deontic status will make the fact that an agent employs, or desires to employ, certain means irrelevant as well. To suppose that the supreme principle of morality makes the issue of whether someone is using another person simply as a means the only relevant consideration with regard to its deontic status is to deny that deontic status is largely objective. This, I will now argue, cannot be correct.

Let us return to the Formula of Humanity.

FH: Act in such a way that you always treat humanity, whether in your own person or in the person of any other, never simply as a means, but always at the same time as an end.[[18]](#endnote-18)

Commentators on Kant generally agree that the reference to ‘humanity’ in FH signifies rational nature. This makes sense when we recall that Kant repeatedly claims that the fundamental principle of morality is binding on all rational beings.[[19]](#endnote-19) Furthermore, the commentators generally agree that the essential feature of rational nature for Kant is the capacity to set ends and judge the desirability of goals for one’s self.[[20]](#endnote-20) Tadros appeals to this idea too.

**Treating Someone as a Means**. We can now examine the significance of treating someone simply as a means. Since Kant incorporates this idea in his formulation of the fundamental principle of morality, we might think that this is the basic condition of moral permissibility. In that case, using people simply as a means is necessary and sufficient for wrongdoing, and not using people simply as a means is necessary and sufficient for permissibility. We can first examine the idea that using people as a means is the basis of deontic status. We can then ask about the significance of the modifier ‘simply’.

Is using a person as a means necessary and sufficient for acting wrongly? Let us put these claims into a revised fundamental principle of morality, FH1:

FH1: An action is not permissible if and only if it treats a person as a means.

This principle entails these two claims:

FH1a. If an action treats a person as a means then it is not permissible;

FH1b. If an action is not permissible then it treats a person as a means. (I. e., if an action does not treat a person as a means then it is permissible.)

For our purposes it is FH1a that is important. If deterrence is an example of treating a person as a means then FH1a tells us that it is not permissible. But we can easily see that FH1a is false. Using people as means is not sufficient for wrongdoing, since we permissibly use people as means in many ways. We permissibly determine what time it is by having others tell us; we are permissibly transported by bus drivers and airline pilots; and so on.

Statement FH1b is also false: using people as means is not a necessary condition of wrongdoing.[[21]](#endnote-21) I will mention four counterexamples.

1. It is false that treating someone as a means is necessary for wrongdoing when harm to her is an unintended but foreseen effect.[[22]](#endnote-22) If S intends to set off a bomb in her backyard for fun, and knows that this will kill her neighbor T, this further effect may be an unintended byproduct of her activity. A fortiori S is not using T as a means. And yet this action is wrong.
2. Harm to a person can be produced as an end, not as a means. S may find it to be desirable as an end to burn down T’s house or to kill T. The fact that doing these things serves no further purpose of S does not mean that they are permissible.
3. A failure to help another person generally is not an example of using her as a means. If S is a Bad Samaritan and walks by T, who is suffering in the road, she does not use T as a means.[[23]](#endnote-23) But S’s action is wrong.
4. Failures to impose obligatory harms such as punishment are problematic. In Kant’s famous example of the ‘desert island’ a governor fails to impose the death penalty on a convicted murderer.[[24]](#endnote-24) If we grant to Kant that this action is wrong, we do not find its wrongness to be due to the fact that the governor is using the prisoner as a means. The governor is not using the prisoner as a means. The point applies more generally. Even if we do not agree that in this case the governor acts wrongly, we do believe that sometimes it is wrong not to impose a punishment. And such inaction is not an example of using the criminal as a means.

In all of these cases the problem stems from insisting that a certain feature of S’s practical

reasoning is a necessary feature of wrongdoing. But S acts wrongly even if this structure is not present.

**Treating Someone Simply as a Means** We can now consider how the word ‘simply’ changes our conclusions.

What would it mean to treat someone simply as a means? I propose to utilize some of Derek Parfit’s suggestions about this.[[25]](#endnote-25) Let us say that S treats T simply as a means if S uses T as a means and S regards T’s interests as of negligible or non-existent moral significance. As Parfit notes, this sort of understanding of the concept means that we cannot determine if S is treating T simply as a means if we only know that S has a certain goal and utilizes certain means in acting with regard to T. We also need to know what S’s attitudes are in general towards T; this would allow us to say how S would treat T in other circumstances. For example, if S tells T the truth only because she expects to be rewarded by T for so speaking, this allows us to infer that S treats T as a means. But we need to know more in order to be able to infer that S is treating T simply as a means. If we know that S would have lied to T if she had been rewarded for doing that, then this suggests that S treated T merely as a means in speaking to her truthfully.

Let us now consider this version of FH:

FH2: An action is not permissible if and only if it treats a person simply as a means.

This statement can be decomposed thus:

FH2a. If an action treats a person simply as a means then it is not permissible;

FH2b. If an action is not permissible then it treats a person simply as a means. (I. e., if an action does not treat a person simply as a means then it is permissible.)

It is only FH2a that is logically relevant to the permissibility of deterrence. If FH2a is true and deterrent punishment uses a person simply as a means, then punishing someone in order to achieve deterrence is impermissible. But because it is enlightening to consider the relevance of an actor’s mental states to the deontic status of her actions, I will briefly consider FH2b.

The change from FH1b to FH2b does nothing to help with the four counterexamples to FH1b. If S treats T simply as a means then she treats T as a means. So if S does not treat T as a means then she does not treat T simply as means. Hence, FH2b also tells us that the actions in the four examples are permissible. These are again implausible claims.

FH2a is more intuitively appealing. There is some plausibility in claiming that if S treats T simply as a means then S acts wrongly. But as Parfit has argued, FH2a is false as well.[[26]](#endnote-26) To see this we can easily modify Kant’s own example. Suppose that merchant S returns the correct change to T as a means of keeping T as a loyal customer and thereby increasing S’s profits. But suppose that S would gladly cheat or even kill T if that would improve her profits. S therefore treats T simply as a means of making money. But it still seems that S acts permissibly in returning the correct change. Or, again, it seems correct to say that S acts in conformity with duty but not from duty.

**Treating Someone as an End**. These reflections suggest that the central idea in FH is really the requirement always to treat humanity as an end.[[27]](#endnote-27) That is, Kant’s view is actually that this is the supreme principle of morality:

FH3: An action is not permissible if and only if it does not treat a person as an end.

It seems like a drastic reinterpretation of FH to say this, given its explicit reference to treating people simply as a means. But I think there is some textual evidence for it.[[28]](#endnote-28) In any case, this interpretation has the virtue of yielding a largely objective conception of deontic status.[[29]](#endnote-29) This does not rule out the possibility that some actions are wrong that involve treating people simply as means. This could be true when treating people simply as means also fails to treat them as ends.

Let us turn to the all-important issue of what it is to treat rational beings as ends. In trying to forestall a natural misunderstanding of his thinking Kant introduces some special terminology. We think of an ‘end’ as a state that an agent seeks to produce by her action. Kant sometimes speaks of this as a “subjective end”.[[30]](#endnote-30) Kant makes a crucial distinction: he says ‘end’ has a different meaning in FH.

Persons…are not merely subjective ends, whose existence as an object of our actions has a value for us: they are objective ends—that is, things whose existence is in itself an end…[[31]](#endnote-31)

Kant thus claims that rational agents are a second sort of ‘end’ that morality properly recognizes. This is a striking and strange use of the term ‘end’.[[32]](#endnote-32) We might say it captures a certain conception of objective normativity, that is, the idea that these ends provide all rational agents with normative reasons to treat them in certain ways. Moral requirements can therefore be thought of as proper responses to the distinctive value that objective ends have. For a given agent S, another agent T constitutes such an objective end. Indeed FH tells us that S is herself an objective end with regard to her own activity.

Kant gives no general characterization in the Groundwork of the value that objective ends have. Commentators who try to expound Kant’s conception of this value therefore use a variety of materials to piece it together.[[33]](#endnote-33) Speaking broadly and roughly, we can say this: treating rational beings as ends, Kant thinks, would consist in not killing or maiming them; not destroying or impairing their rational capacities; not coercing them; and giving them assistance in achieving the ends they rationally set for themselves.[[34]](#endnote-34) In the Groundwork Kant uses FH to establish other obligations, for example, of truthfulness or non-deception, and obligations to refrain from theft, but his reasoning here is, in my opinion, less clear and convincing.[[35]](#endnote-35)

Let us grant, though, that the conception of persons as objective ends can be used to generate a system of obligations and prohibitions. The concepts involved in stating these obligations and prohibitions—deception, coercion, killing, assistance, for example—will be suitably objective in our sense: one can do these things (and refrain from doing them) from many different motives. S can carry out her obligation to assist T in furthering her ends ‘from duty’, but she can also do this from self-interest, and from many other motives. S can likewise carry out her obligation not to kill T from many different motives. And on the other hand, one can perform wrong actions like killing and lying from many different motives. Furthermore, one can perform the relevant actions, whether right or wrong, for their own sakes, or as a means, or by producing the relevant consequences as known but non-intended effects. If S blows up her bomb while foreseeing the death of T, her doing this can be seen as violating FH3: her action destroys a rational being. Again, if S believes that killing T is desirable as an end, and intentionally kills T, this can be seen as violating FH3: S’s action destroys a rational being. Treating people as ends can thus be said to consist in behaving in ways that suitably relate to their moral status as rational goal-setters.

This rough sketch of an interpretation of FH enables us to see how the notion of treating people as ends yields a largely objective conception of the deontic status of actions. This is a reason for supposing that it really is what FH is asserting.

III

Tadros’ Means Principle

My examination of FH was meant to help us think about a central issue in the philosophy of punishment, namely, the permissibility of deterrence. As a useful way of returning to that issue, we can consider how it applies to a moral principle that plays a central role in Tadros’ book, The Ends of Harm.

Tadros accepts the Means Principle, which I suggested can be formulated thus:

MP: If an action harms a person simply as a means of producing good for other persons then it is not permissible. [[36]](#endnote-36)

This principle has some similarities to FH, and Tadros supports MP by appealing to the Kantian idea that “as humans, we are capable of determining what ends to set for ourselves”.[[37]](#endnote-37)

Tadros rejects consequentialism. He asserts that it does not recognize the moral status we have in virtue of our nature as rational goal setters. Consequentialism therefore permits using some persons merely as means of producing good for other persons. He contrasts the well-known examples that he calls ‘Bridge’ and ‘Trolley Driver’. In Bridge a bystander can push a large person off a bridge, which will stop a runaway trolley and save the lives of five people on the track ahead. A consequentialist will say that pushing the large person into the path of the trolley is permissible. Tadros believes that doing this is wrong; however, he believes that—in the contrasting example—turning the trolley away from a track with five people on it onto a track with one person on it is permissible. MP explains why pushing the large person is wrong, while turning the trolley is permissible. In the latter case the single person that the trolley will hit is not harmed as a means, let alone simply as a means.

As I said, Tadros does not believe that MP is absolute; he thinks that there are cases where a person may permissibly be harmed in order to benefit others. For example, if S harms T and refuses to compensate T for her loss, S may be compelled to compensate T. In this case S may permissibly be used simply as a means to benefit T. It is via this sort of exception to MP that Tadros tries to establish, through an extended set of arguments, a right of the state to punish offenders in order to deter potential offenders. The right of the state to do this derives from the right of people to defend themselves, and the duty of wrongdoers to compensate their victims. If wrongdoers refuse to do this, victims have the right to compel them to do so. Such compelled compensation may involve protecting victims from future harm by other wrongdoers. Finally, the state in some circumstances can be regarded as having had this right transferred to it.

Tadros, then, does not object in principle to punishing people in order to achieve deterrence, and he grants that it is permissible in certain circumstances. But Tadros uses MP to object to consequentialism in general, and to its approach to punishment in particular. I will consider whether Tadros can use MP to rule out the pursuit of general deterrence in the design of a system of punishment—unless we can find an exception to MP that somehow allows for it. I believe that our examination of FH can help us to see why it cannot be so used.

Let us begin by considering how MP is related to FH. MP could not conceivably be the fundamental principle of morality, since it gives only a sufficient condition of moral impermissibility. Principles like FH1 and FH2 could conceivably be the fundamental principle of morality, since they give necessary and sufficient conditions of moral impermissibility. Tadros understands this point, since he claims that MP is one of a number of specific moral duties or principles. But there are some other features of MP, besides its logical form, that set it apart from a principle like FH2. First, MP mentions harm; FH2 does not.[[38]](#endnote-38) This makes sense because the fundamental principle of morality must be able to establish that certain kinds of actions which do not harm people—like some types of deception—are wrong. MP need not do this. Second, the mention of ‘harm’ means that MP is not completely subjective—in my sense of the term—in the way that FH1 and FH2 are. MP does not say that only features of a person’s practical reasoning are relevant to this kind of wrongness. Finally, neither FH1 nor FH2 mentions ‘producing good for others’. This seems to mean that FH2, for example, allows for the possibility that it could be wrong for S to use T simply as a means to benefit T. Indeed this may be the reason why Kant stated that “punishment…can never be inflicted merely as a means to promote some other good for the criminal himself.”

I argued that neither FH1 nor FH2 could be regarded as stating plausible necessary and sufficient conditions for wrongness and permissibility. MP only states a sufficient condition for wrongness. Therefore, all of the problems that I noted besetting FH1 and FH2 do not carry over to MP. But since MP is supposed to be a sufficient condition of wrongness it could establish that deterrence is wrong, unless there is an exception to it.

Let us consider whether MP states a plausible sufficient condition for wrongness. And to avoid the problem of the exceptions to that statement that Tadros eventually grants, let us consider fairly ordinary situations that MP does cover. Suppose that S informs the police of the whereabouts of the fugitive T only in order to receive the announced reward, and that S would kidnap or kill T if that were necessary to procure the reward. S brings about harm for T simply as a means of benefitting herself; however, S does not act wrongly. Or suppose that R is a sheriff who evicts tenants from their homes in accordance with legal procedures only in order to support her family, and R would kidnap or kill them if that were necessary to support them. R does not act wrongly in evicting them. MP thus has intuitively implausible implications.

Examples such as the reward-seeker S or sheriff R are more or less parallel to Kant’s self-interested merchant. If we think of the sheriff as morally required—say, by her oath of office—to evict the tenant, then the parallel is exact. The sheriff and the merchant both act ‘in conformity with [moral] duty’. The reward-seeker may have no moral obligation to report the location of the fugitive, but she acts permissibly in doing so. The deontic status of mere permissibility is largely objective, in the same way that wrongness and obligation are. If someone performs an action such as informing the police of the location of T she acts permissibly if she does so in order to benefit herself or her family, and even if she does so with an attitude of complete indifference to T’s well-being. The Kantian distinction between acting in conformity with duty and acting from duty generalizes, mutatis mutandis, to all three deontic categories, and this, too, has intuitive support.

These points have theoretical significance. Kant asserted that his self-interested merchant acts in conformity with duty but not from duty. Kant made this deontic judgment about the merchant and asserted in the same work that FH is the fundamental principle of morality. My two examples give further support to my argument that if we take seriously the idea that moral rules constitute ways of respecting rational goal-setters then we arrive at the idea that FH3 is the best way to formulate FH. FH3 allows us to say that the reward seeker, the sheriff and the merchant all act permissibly. So FH3 does not lead us to suppose that MP is a specific moral rule.

In sum, MP is not a plausible moral principle and it is not a plausible way of specifying how Kant’s idea of respecting rational goal setters applies to the sorts of activities that harm them. It is therefore doubly unsuited to be the way to frame a discussion of whether deterrent punishments are permissible.[[39]](#endnote-39)

My arguments might be thought to establish that MP is more accurately formulated as follows:

MP1: If an action harms another person it is morally impermissible.

However, MP1 is problematic. It seems to rule out the more severe forms of punishment for any purpose whatsoever; it also would mean that the actions of the reward seeker and the sheriff are wrong.[[40]](#endnote-40) Let us go back to the fundamental moral principle in Kant’s philosophy and see what it tells us about the pursuit of deterrence.

IV

Treating People as Ends and Deterrence

Kant asserted that FH is the fundamental principle of morality; Tadros may agree. I argued that FH3 is the most plausible way to interpret FH. Does FH3 permit governments and other institutions to threaten to punish and actually punish offenders in order to deter them and others from wrongdoing? That is, does punishment that is designed to deter potential wrongdoers treat those subjected to it as ends? Put in this form, the question has not been much examined.[[41]](#endnote-41) I will present evidence for saying that FH3 does not support an objection in principle to the pursuit of deterrence.

The most plausible way to frame the discussion is familiar. It involves taking FH in a contractarian direction.[[42]](#endnote-42) If we grant that each rational being is an objective end then we try to find a way to represent the morally significant interests of each of them. We then try to conceive of a sort of moral legislature, where representatives would seek to advance the morally significant interests of the relevant group of rational beings.[[43]](#endnote-43) Treating people as ends would then be conceived of as a derivative phenomenon. It would consist in behaving in ways that conform to the rules that the legislature adopts. These rules can have exceptions. But the rules would be largely objective, in the sense discussed above—that is, parallel in form to MP1—as would exceptions to them. So, for example, if there is an exception to MP1 that permits people to report the whereabouts of fugitives from the law, people may do this simply as a means of supporting themselves or their families, and for various other reasons. The contractarian approach does not rule out the possibility that an agent’s motivation can affect the deontic status of his action, but this is unlikely to occur often.[[44]](#endnote-44) We will see that deterrent punishment is not such a case.

John Rawls’ ‘original position’ is the most well-known philosophical device for conceiving of this legislature, but there are others. Indeed, Rawls himself described a number of possible variants of his own favored description of the legislature. I will examine Rawls’ device partly because it is familiar, and partly because it is well-known that it has features that seem to tilt the deliberation of the legislators away from endorsing utilitarianism and towards his two principles of justice.[[45]](#endnote-45) I also use Rawls because we eventually came to have an excellent study of how Rawls’ approach to justice would apply to the issue of legal punishment, thanks to Sharon Dolovich.[[46]](#endnote-46)

The application of Rawlsian reasoning to punishment is not straightforward. Rawls’ own work after 1955 does not address issues about punishment, even though he focused on the justice of the basic structure of a society’s political and legal system. This curious feature of his work stems from a paradox in A Theory of Justice (as well as the later Political Liberalism): on the one hand, he stipulates that the moral legislators are in, and know that they are in, what he calls the ‘circumstances of justice’.[[47]](#endnote-47) These circumstances make behavior in conformity to the principles of justice, whatever they turn out to be, difficult. On the other hand Rawlsian moral legislators are to select rules that will apply to a situation of ‘strict compliance’. That is, the legislators are to assume that the agents they represent will comply fully with whatever substantive principles of distributive justice they choose.[[48]](#endnote-48) In this sense, the principles chosen are part of ‘ideal theory’.[[49]](#endnote-49)

In characterizing the circumstances of justice Rawls draws on the work of Hart, Hume and (indirectly) Hobbes.[[50]](#endnote-50) These conditions include moderate scarcity of resources, physical vulnerability, limited altruism, and “various shortcomings of knowledge, thought and judgment.”[[51]](#endnote-51) Hart, Hume and Hobbes all envision punishment as necessary for mitigating some of the most problematic consequences of that set of conditions.[[52]](#endnote-52) Yet the principles of justice in Rawls’ theory are to be chosen on the assumption that they will be very largely obeyed with little need or no need for legal punishment.[[53]](#endnote-53)

As Dolovich recognizes we therefore need to change Rawls’ assumptions and goals in order to apply his contractarianism to the problem of legal punishment. The moral legislators have presumably already adopted, at least in outline, the primary substantive moral rules prohibiting acts like murder and theft. But they know that for various reasons some people will not conform to them. In considering whether to use legal punishment they will assume that it may have an effect on how much compliance with the primary rules comes about. That is, they assume that deterrence might affect the rates at which various crimes occur. We can also suppose that rules have already been adopted that establish, at least in outline, a system of governmental institutions involving legislatures, courts and the executive branch; the basic protections afforded by the rule of law; and procedural safeguards for criminal trials.

One feature of the circumstances of justice needs emphasis: imperfections in rationality. As we saw, Rawls takes note of this. It has long been contended that many criminals are imperfectly rational. The imperfections in their rationality could be of various sorts. It is often said to consist in defects in formal rationality, as we might call it, that is, in calculating probabilities, or in discounting the future.But it might also consist in defective ‘weighing’ of moral considerations as against self-interest, or in moral cognition, or in the strength of the agent’s will.None of thisneed mean that all criminals or potential criminals are irrational to such an extent that they should be excused from punishment. But it does mean that the legislators who represent criminals or potential criminals should be more rational than the people they are representing.[[54]](#endnote-54)

We are thus considering this question: would moral legislators, when considering which rules will govern legal institutions operating in the circumstances of justice, allow criminal laws to aim at deterring potential offenders?

Dolovich’s argument for an affirmative answer can be summarized as follows.[[55]](#endnote-55) The two relevant social positions in this question are potential crime victim and punished offender.[[56]](#endnote-56) Behind Rawls’ veil of ignorance the legislators would not know whether they are representing one or the other. (They could be representing both, since one person could be a victim and a punished offender). These legislators also do not know what the probability is of the person whom they represent being a victim or a punished offender.

Dolovich accepts Rawls’ guiding idea that behind the veil of ignorance legislators would adopt a cautious decision-rule, maximin. Maximin requires them to choose principles that maximize the level of well-being of the worst-off representative person. The relevant components of the well-being of the persons being represented in our problem are the security and integrity of the person. These are the main interests protected by the core of the criminal law.[[57]](#endnote-57) Now, if punishments are allowed they would decrease the security and integrity of the persons who are punished. But if deterrence takes place as a result of these punishments then the security and integrity of potential victims of crimes is improved. In some cases the level of security and integrity that offenders arrive after being punished for a given crime will still be higher than the level of the victims of their crimes. These sorts of punishments would be the ones allowed by legislators using maximin. That is because these punishments would raise the level of well-being of the worst-off representative person, the victim, without bringing the level of well-being of punished offenders below it. This claim should be put more precisely, because deterrence is rarely complete, so that the most that is achieved is that there are fewer victims, rather than none. If we think of deterrence as reducing the number of victims then it would be endorsed, Dolovich says, by the legislators in virtue of their use a subsidiary Rawlsian decision-rule, ‘leximin’. Leximin tells its user to maximize the well-being of the worst-off, and then to maximize the level of the next worst-off, and so on.[[58]](#endnote-58)

Dolovich’s argument, in my opinion, is successful, as far as it goes. I would say that it establishes that Kant’s Formula of Humanity, on one plausible interpretation of its application to state punishment, permits the pursuit of deterrence. These further points are worth making:

1. The contractarian reasoning leading to whatever exact principle would be chosen to guide (actual) legislators and other legal officials cannot be seen as disregarding the interests of punished offenders. Their interests are represented in exactly the same way as those of potential victims.
2. Any official who imposed a punishment in conformity with such a principle could therefore be described by a Kantian as treating the offender as an end.
3. There is every reason to think that the principle chosen by the moral legislators will be objective, since it should be comparable to whatever principle governs the sheriff and the informant described above—and, indeed, Kant’s merchant. This would allow us to say that an official like a jailer could permissibly impose a punishment simply as a means of supporting herself or her family. This jailer could correctly be described by a Kantian as acting in conformity with duty but not from duty.[[59]](#endnote-59)

V

One Advantage of Consequentialism

Dolovich’s argument goes some way to establishing that there is no difference with regard to the permissibility of deterrence between a Kantianism based on the Formula of Humanity and consequentialism. Kantianism (plausibly interpreted) and consequentialism allow the pursuit of deterrence. In fact, consequentialism and Kantianism share some fundamental similarities. Each theory weighs the interests of victims and potential criminals in the same way. And both theories make deontic status largely, if not entirely, objective, so that it is possible for someone to act rightly even if she is using another person simply as a means of benefitting herself, or someone else, or, indeed, society as a whole.

Having established these similarities, I want to explore briefly how they differ. In particular, I will consider what upper limit exists on the severity of punishments designed to achieve deterrence, using Dolovich’s work as a point of departure. Her Kantian approach to the limits on deterrence is less convincing than that of consequentialism.

Dolovich’s main conclusion about the upper limit on severity is basically this: the loss imposed on an offender by the punishment for a crime may not exceed the loss suffered by the typical victim of it.[[60]](#endnote-60) That is, the loss suffered by a potential individual victim constitutes a sort of ceiling on the severity of acceptable punishments.[[61]](#endnote-61) Dolovich grants some exceptions to this basic principle[[62]](#endnote-62), but I will not focus on the circumstances where these are operative.

The argument for this sort of ceiling is as follows. The two basic relevant social positions are punished offender and potential victim. Behind the veil of ignorance legislators know only the levels of well-being of persons in those positions. Using maximin (or leximin) they will seek to make the position of the worst-off representative person as high as possible. The legislators will allow deterrent punishments to bring down the level of well-being of offenders. But they will not allow these sanctions to be so severe that punished offenders end up in a worse position than potential victims.

Consequentialism is often supported by arguments that involve appealing to the impartial consideration of interests with full information. If this is done when the severity of punishment is the issue we would not be limited to considering the levels of well-being of a punished offender and a potential victim. We also would have information about the probability that an offender would be punished for a certain type of crime, and the probability that a person would be victimized in that way.

There is a well-known problem with approaches to punishment that fail to take this sort of information into account. Certain crimes like burglary are rarely punished in a society like the US, since the chances of being apprehended are low. Dolovich’s Kantian approach would limit the punishment for one act of burglary to the loss suffered by a victim of this crime—presumably the typical one. But since many burglars often engage in that crime, and are rarely punished, such a limited punishment will have little deterrent effect.[[63]](#endnote-63) For example, suppose that a burglar has a 1 in 100 chance of being punished, and the typical victim suffers a loss of, say, $1000 and also experiences the fear resulting from having been victimized. The punishment for this action, in Dolovich’s approach, has to be limited to something with this level of severity. (The victim will presumably also have the right to sue for compensation.) A consequentialism that allowed probabilities to be reckoned on at the fundamental level in designing punishments would not be so constrained. That is, it would allow the punishment of a burglar to be more severe than the loss of $1000 plus whatever corresponds to the fear the homeowner experiences. This is a point in favor of consequentialism.

I am not claiming that the consequentialist approach to the limits on severity is therefore convincing in all cases. But I do claim that on this type of problem it seems to be better than the version of Kantianism presented by Dolovich.

VI

Conclusion

The assertion that we may not use people simply as a means has suggested to many philosophers that the pursuit of deterrence is morally objectionable in principle. And the conviction that we may not harm people simply as a means led Victor Tadros to look for a justification of punishment that does not rest on a consequentialist basis. I argued that the Formula of Humanity is not plausibly construed as containing the ‘simply as a means’ phraseology. The more plausible understanding of it requires us always to treat people as ends. I then argued that this formula permits the pursuit of deterrence. The best understanding of Kantian moral theory does not conflict with consequentialism on this fundamental point in the philosophy of punishment. Finally, I gave one reason for thinking that consequentialism handles one further issue about the permissible severity of deterrent punishments better than a version of Kantianism based on the requirement to treat people as ends.[[64]](#endnote-64)

1. For an explanation of the error, see my “Punishment and Reform” Criminal Law and Philosophy (forthcoming). [↑](#endnote-ref-1)
2. In “The Origins of the ‘Objection,’” History of Philosophy Quarterly 29 (2012), 79-101, I show that this objection to utilitarianism does not occur in Kant, and was developed much later. [↑](#endnote-ref-2)
3. Immanuel Kant, Groundwork of the Metaphysic of Morals. Trans. H. J. Paton. (New York: Harper and Row, 1964), 96 (= 429 in the Prussian Academy Edition). This work is hereafter abbreviated ‘G’. Emphasis deleted. [↑](#endnote-ref-3)
4. Immanuel Kant, The Metaphysics of Morals. Trans. Mary Gregor (Cambridge, Eng.: Cambridge University Press, 1991), 140-1 (= 331 of the Academy edition). Hereafter, ‘MM’. [↑](#endnote-ref-4)
5. It also seems to undermine the interpretations of scholars who contend that the Doctrine of Right in MM is not an example of an “applied ethics” where the Categorical Imperative (and in particular, FH) is directed to issues in political and legal philosophy. See Arthur Ripstein, Force and Freedom (Cambridge, Mass.: Harvard University Press, 2009), 6, 11-12; Allen Wood, “The Final Form of Kant’s Practical Philosophy.” Kant’s Metaphysics of Morals, ed. Mark Timmons. (Oxford: Oxford University Press, 2002), 5-10. [↑](#endnote-ref-5)
6. See the authoritative article by Sharon Byrd, “Kant’s Theory of Punishment,” Law and Philosophy 8 (1989), 151-200. Cp. B. Sharon Byrd and Joachim Hruschka, Kant’s Doctrine of Right: A Commentary (Cambridge: Cambridge University Press, 2010), ch. 13. [↑](#endnote-ref-6)
7. R. A. Duff, Trials and Punishments (Cambridge, Eng.: Cambridge University Press, 1986), 178-86; R. A. Duff, Punishment, Communication, and Community (Oxford: Oxford University Press, 2001), 82-5; Andrew von Hirsch, Censure and Sanctions (Oxford: Oxford University Press, 1993), 11-13; Jean Hampton, “The Moral Education Theory of Punishment”. Philosophy and Public Affairs 13 (1984), 208-38, at 211f. Duff, von Hirsch and Hampton appeal to Hegel’s claim that deterrent threats treat rational agents “like a dog”. G. W. F. Hegel, Elements of the Philosophy of Right. Trans. H. B. Nisbet (Cambridge, Eng.: Cambridge University Press, 1991), 125. The focus of Kant’s remarks in MM is the act of imposing of punishment; Hegel’s is on the threat of its imposition. But both are asserting that features of deterrence constitute a deep affront to an agent’s status as a rational being. [↑](#endnote-ref-7)
8. Robert Nozick, Philosophical Explanations (Cambridge, Mass: Harvard University Press, 1981), 363-97, esp. 372; Michael Moore, Placing Blame (Oxford: Oxford University Press, 1997), 27-30, 153-88. [↑](#endnote-ref-8)
9. Cp. Victor Tadros, The Ends of Harm (Oxford: Oxford University Press, 2011), 114. Hereafter, ‘EH’. Tadros himself usually states MP as prohibiting harming a person as a means of producing good for others. The bracketed word is often not stated by Tadros, but it is clear that he accepts part of the idea that it is usually understood to convey. See EH 135-6, where he grants that if a person consents to be harmed for the benefit of others then MP is not violated, since, he says, in that case she is not used “merely” for the good of another. My discussion in Section II discusses the distinction between using someone as a means and using her simply as means. In Section III I interpret Tadros as making this distinction as well. Doing so is being charitable, I think, since it makes MP less vulnerable to counterexample. [↑](#endnote-ref-9)
10. EH 127. On Tadros’ general debt to Kant, see EH 2. For Kant’s emphasis on the ability to set goals see G 105 (437); MM 195 (392). On the other hand, Tadros sometimes implies that his Means Principle is a variant of the Doctrine of Double Effect, which has a rather different intellectual provenance and basis. EH 116; 150. The Doctrine of Double Effect clearly distinguishes between foreseen and intended effects, but FH is not plausibly seen as doing this—or so I will argue in Section II. Hence, the rule about harming that would most straightforwardly be derivable from FH would not be MP. See also n. 44 below. [↑](#endnote-ref-10)
11. Steven Sverdlik, Motive and Rightness (Oxford: Oxford University Press, 2011). See pp. 103-16. Hereafter, ‘MR’. [↑](#endnote-ref-11)
12. The best general discussions of FH are Thomas Hill, Dignity and Practical Reason (Ithaca: Cornell University Press, 1992), ch. 2 (hereafter, ‘DPR; Allen Wood, Kant’s Ethical Thought (Cambridge, Eng.: Cambridge University Press, 1999), ch. 4. Hereafter, ‘KET’. For their conclusions about the ‘simply as a means’ phraseology see DPR 41-2; KET 142-4. [↑](#endnote-ref-12)
13. G 65 (397). [↑](#endnote-ref-13)
14. See MR, chapter 2. [↑](#endnote-ref-14)
15. John Stuart Mill, Utilitarianism. Ed. George Sher (Indianapolis, IN: Hackett, 1979), 17-18. Hereafter, ‘U’. [↑](#endnote-ref-15)
16. See MR, especially ch. 4. For the general question see also Thomas Scanlon, Moral Dimensions (Cambridge, Mass.: Harvard University Press, 2008), chs. 1-3; EH, ch. 7. [↑](#endnote-ref-16)
17. U 17. See also MR chs. 3-4; Steven Sverdlik, “Consequentialism, Moral Motivation, and the Deontic Relevance of Motives.” Moral Motivation, ed. Iakovos Vasiliou. (Oxford University Press, forthcoming). [↑](#endnote-ref-17)
18. G 96 (429). Emphasis deleted. [↑](#endnote-ref-18)
19. G 57 (389); 92 (425). [↑](#endnote-ref-19)
20. G 105 (437); MM 195 (392). See DPR 38-41; KET 118-22. For dissent: Richard Dean, The Value of Humanity in Kant’s Moral Theory (Oxford: Oxford University Press, 2006), chs 2-4. [↑](#endnote-ref-20)
21. MR 107-8. [↑](#endnote-ref-21)
22. See Jonathan Bennett, The Act Itself (Oxford: Oxford University Press, 1995), p. 218. [↑](#endnote-ref-22)
23. DPR 42. [↑](#endnote-ref-23)
24. MM 142 (333). [↑](#endnote-ref-24)
25. Derek Parfit, On What Matters (Oxford: Oxford University Press, 2011). Vol. I, ch. 9, sec. 30, pp. 212-32, at 213-5. Hereafter, ‘OWM’. Hill and Wood hardly consider this question, presumably because they find the ‘simply as a means’ phraseology unnecessary. DPR 41-2; KET 142-3. [↑](#endnote-ref-25)
26. OWM 216. [↑](#endnote-ref-26)
27. Wood reaches this conclusion by directly examining its wording. KET 143. Cp. DPR 41-2. [↑](#endnote-ref-27)
28. See MR, 110, n. 21. [↑](#endnote-ref-28)
29. It is true that Kant himself veers away from objectivism in MM. See MR 117-25. His claims in MM would need to be rethought, if FH3 is the correct way to interpret FH. [↑](#endnote-ref-29)
30. E.g., G 96 (428). [↑](#endnote-ref-30)
31. G 96 (428). Emphasis deleted. “Objective end” also occurs at G 98 (431). Cp. G 96 (429): “an objective principle”. Cp. also G 105 (437). [↑](#endnote-ref-31)
32. Henry Sidgwick calls it “perplexing”. The Methods of Ethics, 7th ed. (Indianapolis: Hackett, 1981), 390. [↑](#endnote-ref-32)
33. DPR 43-50, draws on a number of passages to provide “clues regarding what it means to treat humanity as an end.” (46) [↑](#endnote-ref-33)
34. For more detail on these claims see MR 111-14. [↑](#endnote-ref-34)
35. MR 114 n. 40. [↑](#endnote-ref-35)
36. Cp. EH 114. For this formulation see n. 9 above. [↑](#endnote-ref-36)
37. EH 127. Cp. 153. The formulation on 124—we are not “available for use by others to promote the good”—is tendentious. ‘Available’ must mean ‘permissibly available’. [↑](#endnote-ref-37)
38. For some other principles prohibiting harming people simply as a means see OWM 228-32. [↑](#endnote-ref-38)
39. This criticism applies a fortiori to the actual formulation that Tadros uses. EH 114. [↑](#endnote-ref-39)
40. It might be said that reformulating MP as MP1 poses no problem for Tadros. If he has presented a set of arguments that carves an exception out of MP then they could also carve an exception out of MP1. So he could proceed to defend the permissibility of deterrence in his non-consequentialist way. But accepting MP1 would create a serious dialectical problem for Tadros. If he accepts MP1 he loses his basis for objecting to consequentialism. MP can be used to show that pushing the large person is wrong but turning the trolley is not. MP1, however, entails that both actions are wrong. [↑](#endnote-ref-40)
41. One noteworthy example is Thomas Hill, “Treating Criminals as Ends in Themselves.” Virtue, Rules and Justice (Oxford: Oxford University Press, 2012), 296-319, esp. 310-11. [↑](#endnote-ref-41)
42. And, perhaps, supplementing FH with the Formula of the Kingdom of Ends. G 439 (106). [↑](#endnote-ref-42)
43. Cp. Hill, “Treating Criminals”, op. cit., p. 304 f. [↑](#endnote-ref-43)
44. See Parfit’s discussion of ‘Tunnel’ (i.e., Tadros’ ‘Trolley Driver’) and ‘Bridge’ in OWM 218-21. One of his points is this: a rational agent would have no preference for being killed as a side effect of the action of turning the trolley in Tunnel as compared to being killed as a means in Bridge. This fact helps to establish that the Kantian Consent Principle that Parfit is discussing would not condemn the act of turning the trolley or the act of putting the person on the bridge in front of the trolley. Hence, Kantian reasoning is not going to generally endorse moral principles like the Doctrine of Double Effect or Tadros’ MP. [↑](#endnote-ref-44)
45. Especially important and controversial is the ‘thickness’ of the ‘veil of ignorance’. Rawls acknowledges that if the legislators know more about the society that they are choosing principles for, some form of utilitarianism would be likely to be chosen instead of Rawls’ conception of justice. See Holly Smith Goldman, “Rawls and Utilitarianism.” John Rawls’ Theory of Justice: An Introduction. Ed. H. Gene Blocker and Elizabeth Smith (Athens, Ohio: Ohio University Press, 1980), 346-94, at 373-5. Cp. 384-8. [↑](#endnote-ref-45)
46. Sharon Dolovich, “Legitimate Punishment in Liberal Democracy.” Buffalo Criminal Law Review 7 (2004), 307-442. Hereafter, LP. Rawls’ A Theory of Justice appeared in 1971. [↑](#endnote-ref-46)
47. John Rawls, A Theory of Justice, 2nd ed. (Cambridge, Mass.: Harvard University Press, 1999), 109-12. Hereafter, ‘TJ’. In Political Liberalism the description of the circumstances of justice becomes considerably simplified, and Rawls’ attention focuses on a new feature, “the fact of pluralism”. John Rawls, Political Liberalism. (New York: Columbia University Press, 1996), 66. Hereafter, ‘PL’. [↑](#endnote-ref-47)
48. TJ 125-6. [↑](#endnote-ref-48)
49. TJ 7-8; 215-6. [↑](#endnote-ref-49)
50. For the Hume and Hart references, see TJ 109, n. 3. But Hart (who frames his discussion as “the minimum content of natural law”) in turn cites Hobbes. See H. L. A. Hart, The Concept of Law, 2nd ed. (Oxford: Oxford University Press, 1994), 189-95, and endnote at 303. [↑](#endnote-ref-50)
51. TJ, 110. [↑](#endnote-ref-51)
52. Hart emphasizes the need for sanctions to enforce basic moral norms at Concept, 193. Hobbes is well-known for doing so. Hume is a more interesting case. Rawls cites only section 2 of Treatise III ii. Hume there describes some of the factors that Rawls mentions, such as moderate scarcity, and explains how the conventional rules of justice arose to ameliorate the difficulties human society faced in its ‘natural condition’. Hume does not mention any enforcement mechanisms, except the loss of reputation. Hart, however, cites sections 4-7 in addition to 2. In section 7, Hume clearly envisions the use of sanctions to combat the effects of narrow passions. David Hume, A Treatise of Human Nature, 2nd ed. Ed. P. H. Nidditch (Oxford: Oxford University Press, 1978), 535. Cp. Hart, Concept, 193. Rawls’ neglect of the role of punishment in his treatment of justice was not due to his misinterpreting his sources, however, but to his adaptation in moral theory of a method used in economics, viz., developing an idealized model as a point of departure for understanding more complicated cases. [↑](#endnote-ref-52)
53. Rawls briefly discusses punishment at TJ 211-2. He grants that even in ideal circumstances some punishments may at least need to threatened, although they may never need to be imposed. [↑](#endnote-ref-53)
54. In TJ Rawls tends to take the parties to be actual citizens, required to reason in a certain way. TJ 109-12. In PL they are described as ‘representatives’ of citizens. PL 24-5; 306-7. Dolovich describes most potential criminals as participating in the deliberations, but not all. LP 358-63. [↑](#endnote-ref-54)
55. Dolovich thinks that the argument she presents would justify specific deterrence, general deterrence and incapacitation. LP 383. She does not consider whether it would justify reformative punishment. [↑](#endnote-ref-55)
56. LP 357. Later Dolovich recognizes a third position, a falsely convicted agent. LP 378. [↑](#endnote-ref-56)
57. LP 354-5. [↑](#endnote-ref-57)
58. TJ 72 (where it is called the ‘lexical difference principle’). LP 379-85. Dolovich assumes that decreasing the number of victims, when punished criminals are not brought down to their level of well-being, is mandated by leximin. This does not seem correct. However, a related principle discussed by Thomas Scanlon, suitably modified, would mandate decreasing the number of victims, when punished criminals are not brought down to the level of well-being of the remaining victims. Thomas Scanlon, “Rawls’ Theory of Justice.” Reading Rawls, Ed. Norman Daniels (New York: Basic Books, 1975), 169-205, at 197. Scanlon’s principle in effect mandates decreasing the probability that a person will be a victim. This could be adopted behind the veil of ignorance, since it could be accepted without knowing what the probability of being a victim is. [↑](#endnote-ref-58)
59. Cp. MR 138-40. [↑](#endnote-ref-59)
60. LP 386-400. Dolovich makes her argument using simplified categories of offenses (‘serious’ and ‘non-serious’), and punishments (humane incarceration and less burdensome impositions). See 386-7. [↑](#endnote-ref-60)
61. This limitation resembles the one in W. D. Ross, The Right and the Good (Oxford: Oxford University Press, 1930), 56-64. But Ross thinks of the limit as established by the loss of the victim of a crime; Dolovich is using the loss of a potential victim. And she recognizes some exceptions to this limitation. See next note. [↑](#endnote-ref-61)
62. See especially LP 399-400. [↑](#endnote-ref-62)
63. See Alan Goldman, “The Paradox of Punishment.” Philosophy and Public Affairs 9 (1972), 42-58. Goldman takes the issue to be an inescapable dilemma for a Rossian approach, which he regards as the best available. I take it, rather, as an objection to it and the similar limitation in Dolovich. [↑](#endnote-ref-63)
64. I am indebted to John Deigh, Robert Howell, Luke Robinson, Justin Fisher, and Tim Henning for specific comments and suggestions. Christian Seidel and the other participants at the 2013 Erlangen consequentialism conference were also very helpful and supportive. [↑](#endnote-ref-64)