

Must Punishment Be Intended to Cause Suffering?

Bill Wringer

Accepted: 24 September 2012 / Published online: 11 October 2012
© Springer Science+Business Media Dordrecht 2012

Abstract It has recently been suggested that the fact that punishment involves an intention to cause suffering undermines expressive justifications of punishment. I argue that while punishment must involve harsh treatment, harsh treatment need not involve an intention to cause suffering. Expressivists should adopt this conception of harsh treatment.

Keywords Punishment · Expressive theories of punishment

1 Punishment and the Intention to Cause Suffering

Philosophical discussions of punishment often start from definitional considerations, such as Hart's well-known characterization of punishment as involving harsh treatment, inflicted on an offender by an appropriate authority, in response to some specifiable form of wrongdoing (Hart 1959). David Boonin (Boonin 2008) and Nathan Hanna (Hanna 2008, 2009a,b) have recently argued that this characterization of punishment either obscures or leaves out something important, namely that when we punish other people, we intend to make them suffer.¹ Boonin and Hanna both think that this fact presents a problem for philosophers who think that the practice of punishing offenders is justifiable. Boonin claims that the idea that punishment involves an intention to cause harm presents a problem for any attempt to justify

¹Some exemplary quotations: 'Punishment involves *acting with the intention of harming someone* because she has (or is at least believed to have) committed an offense'. Boonin 2008 p 21 (italics mine).'; 'To punish an offender, one must aim to hurt or harm her in some way; the treatment to which she is subjected must be used at least in part as a way of making her suffer.' Hanna 2008 p 126.

B. Wringer (✉)
Department of Philosophy, Bilkent University, Ankara, Turkey
e-mail: wringer@bilkent.edu.tr

punishment. Hanna argues that it plays a significant role in undermining one particular strategy for justifying punishment, which he calls expressive retributivism.

Here I shall argue that the claim that punishment must involve an intention to cause suffering is mistaken. Since punishment need not involve an intention to cause suffering, attempts to show that punishment is unjustified which rely on this claim cannot be sustained. My case draws on ideas developed within the expressivist retributivist tradition that Hanna criticizes. I shall argue that if we take expressive goals of the right sort to form part of the definition of punishment, treatment can be characterized as harsh even if it is not intended to cause suffering—something which Hanna explicitly denies.

If punishment must involve an intention to cause suffering, attempts at punishment which do not make offenders suffer are failures. It might then seem appropriate to object to a punitive regime visited on some individual that it does not make them suffer (or that it does not make them suffer enough), or to attempt to assess the degree to which a convicted offender has actually suffered, and to supplement the punishment if they have not. However, if punishment need not involve an intention to cause suffering, this need not be appropriate.²

2 The Importance of Intentions

Hanna and Boonin's objection to punishment is precisely that it involves an *intention* to cause suffering, and not for example, that it does so reliably or foreseeably. Here is what Hanna says:

In what follows I will assume that the aim to impose suffering is essential to punishment, and will argue that this characteristic, in conjunction with other considerations undermines ER.³
(Hanna 2008 pp 127–8)

The 'other considerations' that Hanna refers to are that there are ways of achieving the valuable goals which the expressive retributivist takes punishment to achieve and which do not involve an intention to cause suffering; (Hanna 2008 pp 137–9) and that there are moral reasons against doing something which is intended to cause suffering to another person in order to achieve a morally valuable goal if there are alternative ways of achieving this goal which avoid doing anything which is intended to cause suffering.⁴

Neither Hanna nor Boonin think that the fact that a practice causes suffering, or even that it causes suffering which is foreseeable makes that practice impermissible in the same way that the intention to cause suffering does. Both hold that requiring criminals to pay restitution is legitimate even when this will harm the criminal, because it does not involve an intention to harm those individuals (Boonin 2008 pp 213–276 Hanna 2008 pp 138–9).

² Of course there may be other ways of blocking this implication.

³ ER is Hanna's abbreviation of 'Expressive Retributivism', a view I sketch below.

⁴ The existence of alternatives to punishment which do not involve an intention to cause suffering only counts against expressivism if this is true. While Hanna 2008 leaves this implicit, Hanna 2009a explains why a liberal might find the intention to cause suffering problematic.

3 Defending Punishment: An Expressivist Response

My response to Hanna and Boonin's critique of punishment draws on the expressive retributivist tradition that Hanna targets.⁵ On an expressive theory of punishment, an important part of the justification of punishment consists in the messages it conveys. Some expressivists follow Joel Feinberg in taking the expressive dimension of punishment to be partially constitutive of something's being punishment. (Duff 2009; Feinberg 1970)⁶ For them, harsh treatment which is a response to wrong-doing by an appropriate authority should only be counted as punishment if it is intended to express a message of disapproval to those who perpetrate those crimes, or to society as a whole, or both. Other expressivists do not take the expressive dimension of a practice of responding to wrong-doing to be part of what makes it into punishment. However, authors of both sorts hold that, whether or not it is constitutive of something's being punishment, the expressive message which punishment is intended to express play a significant role in justifying punishment.

Different expressivists have different views about the role the expressive dimension of punishment plays in justifying punishment. On a 'denunciatory' view of punishment, the purpose of the harsh treatment that punishment involves is for a society to communicate to its members that certain norms are in force and that transgressions against them are viewed seriously. (Metz 2000; Wringle 2006) On Duff's 'communicative' view, punishment is a way in which a society can communicate its condemnation of wrongdoing to offenders in ways which prompt them to experience regret and remorse, both of which can play a valuable role in allowing them to be re-integrated into society. (Duff 2001). For Christopher Bennett, punishment is a way in which society expresses certain 'reactive attitudes'. Punishment's justifiability derives from the right criminals have to be punished, which derives from their right to be treated as fully responsible members of their political communities, and to undergo a form of treatment which involves a form of symbolic penance. (Bennett 2008)

Expressivists often suppose that the state has a duty to communicate certain messages about certain kinds of wrong-doing either to the individuals who have acted wrongly (as on Duff's view) or to society as a whole (as on Metz's).⁷ This point is not enough to justify imposing on offenders the sorts of treatment which we typically take punishment to involve: imprisonment, community service, monetary fines and the like—forms of treatment often involve things which those subjected to them would rather not undergo. Call treatment of this sort 'burdensome treatment'. (On this definition, whether a form of treatment is burdensome depends on how those on whom it is inflicted respond to it, and not on the intentions of those who impose it. I may impose treatment which I know to be burdensome on someone with the intention of making them suffer. But I may also impose burdensome treatment without having

⁵ I shall generally speak of 'expressivism' where Hanna talks of 'expressive retributivism' (though many expressivists are also retributivists). Representatives of expressivism differ on many issues: I shall discuss versions defended in Duff 2001; Bennett 2008 and Wringle 2006. One referee doubts whether Duff is an expressivist. However, like Hanna (Hanna 2008 p 142) and Boonin (Boonin 2008 p 172), I regard him as an particularly important representative of the view. Other important contributions to expressivism include Hampton 1992 and Metz 2000, who defends a view that has something in common with Wringle 2006: (I thank a referee for this journal for drawing this paper to my attention).

⁶ Duff sometimes suggests that this is part of the definition of punishment, and sometimes that it is central to the justification of punishment, but not definitional. See eg Duff 2001 pp xiv–x.

⁷ Compare Metz 2000

this intention: for example, I may think that someone is so hardened, or so desperate that a short stay in prison will do them no harm, but choose to impose it anyway. It be that I am mistaken in this belief. If so, I will have imposed burdensome treatment without any intention to cause suffering).

Different expressivists explain why we need burdensome treatment to achieve the communicative goals of punishment in various different ways. For example, Duff suggests that in punishing an individual, we are communicating something to him or her about the wrongness of his or her action in the hope that this will bring about repentance and remorse. Imposing burdensome treatment is a way of trying to make the wrongdoer focus on their wrongful action in a way which they might be less likely to do if the state communicated with them about the wrongfulness of their acts in some other way, such as, sending a condemnatory letter from a magistrate or other state official. (Duff 2001 pp 107–9) On Wringer's view the imposition of burdensome treatment is a way in which a punishing authority can express to a community that the infringement of certain kinds of norms is taken seriously: again, the suggestion is that purely symbolic, non-burdensome acts of condemnation could not do this as effectively. (Wringer 2006 p 180) Finally, Bennett argues that the imposition of something burdensome is a way in which the state can express a form of censure which is 'symbolically adequate' to the offense: because it does so, the offender is compelled, as Bennett puts it 'to undertake proportionate apologetic action as a way of undoing public wrong.' According to Bennett the censure must involve something burdensome, in order to be 'symbolically adequate': what the punished individual undergoes should be able to be understood by society at large as an appropriate and proportionate form of apology; and only burdensome treatment can be so understood. (Bennett 2008 pp 106–7).

If an expressivist conception of punishment is correct, punishing someone need not involve an intention to cause suffering. Or so, at any rate, I shall claim. I shall argue that although punishment must involve harsh treatment, it does not follow from a form of treatment's being harsh that imposing it must involve the intentional imposition of suffering. Furthermore treatment which is harsh but which does not involve an intention to cause suffering can fulfill the expressive goals which expressivists take to play an essential role in justifying punishment.

4 Understanding Harshness

Why do Hanna and Boonin hold that punishing someone must involve an intention to cause them to suffer? This claim seems to go beyond the characterization of punishment given in the first paragraph of this paper. On the face of it, that characterization makes no reference to such an intention: it only speaks of the treatment involved being 'harsh'. The same is true if we add to this minimal characterization the requirement, drawn from Feinberg, the expressivist claim that the harsh treatment be at conveying a message of disapproval, either to the offender, or to society as a whole.

The minimal characterization does make some reference to intentions, since it says that something can only be punishment if it is a response to wrongdoing; and, plausibly, the phrase 'a response to wrongdoing' is best understood in a way which presupposes that something can only constitute a response to wrongdoing if those who are inflicting it intend what they are doing as a response to wrongdoing. It is also implausible that harsh treatment inflicted on a wrongdoer *accidentally* constitutes punishment, even if it could be correctly

described as a response to wrongdoing. Nevertheless, having an intention to respond to wrongdoing need not entail having an intention to cause suffering.

Hanna argues that something important is obscured by the fact that the characterization of punishment that I have given in the first paragraph makes no explicit reference to punishment's involving an intention to cause suffering. On his view, the notion of harsh treatment is best understood as treatment that it is intended to cause suffering. (Hanna 2008 pp 124–7; 2009a p 330). I disagree: I shall argue that the right way of understanding the notion of harsh treatment which figures in the characterization of punishment given in the first paragraph of the paper is that it is treatment which would normally be found burdensome by a typical individual of the kind on whom it is being imposed. Notice that while it is true that, on an account of this sort, punishment need not involve an intention to cause suffering, it is nevertheless a non-accidental feature of punishment that it often does cause suffering. It is thus consistent with a claim Duff makes and Hanna highlights:

Punishment aims to inflict something painful or burdensome on an offender ... nor are this pain and this burden mere unintended side effects of a procedure which is not designed to be painful or burdensome.

(Duff 1992 p 49, Hanna 2008 p 127)

My case against Hanna does not turn on the verbal difference between something's being burdensome, and something's causing suffering, but the difference between something's being actually intended to cause suffering to a particular individual, and its merely being of a kind which is liable to cause suffering to individuals of that kind.

The existence of a difference is established by the following argument. If I inflict on you some treatment which is designed to cause you to suffer, and you do not in fact suffer, then I have to that extent failed. If my goal is important enough, it may be appropriate to inflict some other treatment on you. But if I intend to inflict on you some treatment which would typically cause suffering in individuals like you, and you do not suffer then I have not necessarily failed. I *may* have failed: I may be mistaken about what individuals like you would find burdensome. Furthermore, if I have not made you suffer, this may be evidence that I am mistaken about exactly that. However, this evidence might be misleading. Perhaps I fail because, very unusually, you have become a convinced Stoic while awaiting sentence, and now believe that imprisonment is not a way in which you can be harmed.⁸ Or perhaps you have fallen into an uncharacteristic state of insensibility. In these cases, I have not necessarily failed in the goal of inflicting treatment that someone like you would normally find burdensome.

In arguing that this conception of harsh treatment is distinct from Hanna's I am appealing to the principles that action types with distinct success conditions are distinct; and that success conditions are distinct if one can be fulfilled while the other is not. These principles are both general and plausible. So the characterization of harsh treatment offered here is distinct from Hanna's. On the other hand it is a characterization on which punishment typically will cause suffering, and on which the fact that it does so is non-accidental. In the normal course of events, not every individual on whom punishment is inflicted will be an atypical individual of their kind.

Admittedly, there is no logical bar to its being the case that every individual on whom punishment is inflicted should be an atypical individual of their kind. However, it is

⁸ Ancient Stoics like Epictetus held that the only way I can be harmed is to become less virtuous. (Epictetus 1928 *passim*). My argument does not require that the reader accept the Stoic view, but only that someone else might do so.

vanishingly unlikely that this will actually occur: it would require a kind of cosmic coincidence. It is not an important objection to an account of harshness that if such a cosmic coincidence were to occur, it would be untrue that harsh treatment typically caused suffering. For we have no need of a justification of punishment that would hold in such unusual cases. An account which shows punishment to be justified in worlds like the actual world, in which there are no cosmic coincidences of this sort should be good enough for us.

One might object that although on my account punishing someone need not involve an intention to cause them to suffer, it involves something equally problematic: namely, the imposition of burdensome treatment on individuals, simply because it is burdensome. Here, we need to make a distinction. We might mean understand 'burdensome' in a way in which it follows from the fact that I have imposed burdensome treatment on X that X has actually been burdened. On my account of harshness, it does not follow from the fact that X has been punished that they have actually been burdened. So in this sense of 'burdensome,' it is not true that on my account punishment imposes anything at all on people *because* it is burdensome. Punishment does not fail of its object if it fails to impose something which is burdensome in this sense.

Alternatively, we might understand 'burdensome' to mean 'that which many people might find to impose a burden.' If this is what 'burdensome' means, my account does entail that punishment involves the imposition of something burdensome because it is burdensome. However, it does not entail that that punishment involves the imposition of something burdensome because it is burdensome *and for no other reason*. Part of the point of the expressive account is that punishment involves the imposition of burdensome treatment because of the expressive goals that this is supposed to achieve. Furthermore, as I pointed out in section II, on some versions of expressivism, *only* burdensome treatment could achieve these goals.⁹ (By contrast, it is not true that only treatment intended to cause suffering could achieve these goals).

There might be something wrong with imposing burdensome treatment on people because it is burdensome and for no other reason. On my account, punishing someone does not do this. One might also think that there is something wrong with imposing burdensome treatment on people because it is burdensome in pursuit of some other worthwhile goal, and in particular in pursuit of a goal to which those on whom the treatment is may not consent. But neither Hanna nor Boonin commit themselves to anything so subtle: what they object to is the intentional inflicting of suffering.

Furthermore, we have already seen that Hanna and Boonin hold that although it is wrong to intentionally inflict suffering on other people, it is not wrong to inflict treatment on people which we can foresee would make them suffer, in order to achieve valuable goals to which they may not consent (namely, the compensation of victims of wrongdoing). It is hard to see how one could coherently regard this as acceptable, and yet regard it as unacceptable to inflict on the same individuals treatment which we can foresee would make other people suffer, if it were imposed on them, in order to achieve goals to which these individuals might not consent. Surely, the only way in which the fact that a form of treatment would foreseeably cause other individuals like the offender to suffer could be morally relevant is that the fact that it would cause those other individuals to suffer is very good evidence for the claim

⁹ We might think that a form of treatment's being burdensome means there is (at least *prima facie*) a *reason* for not imposing it, even if it does not follow that we have a duty not to impose it. But this point would not count very strongly against versions of the expressive view which hold (as Bennett and Metz do) that societies have a duty to punish. For this duty would presumably outweigh these reasons against imposing burdensome treatment.

that this treatment would foreseeably make the offender suffer. If this can, in fact, be morally acceptable—as Boonin and Hanna think—then it seems as though the intentional inflicting of burdensome treatment on individuals to achieve valuable goals can also be acceptable.

5 Quarantine and the Involuntary Detainment of the Mentally Ill

Hanna argues that the existence of cases of burdensome treatment which are not punishment, such as quarantine and involuntary detainment of the mentally ill, shows that in characterizing punishment as harsh treatment, we think of harshness as involving an intention to cause suffering. Quarantine and the involuntary hospitalization of the mentally ill are not punishment because they do not involve an intention to cause suffering. (Hanna 2008 pp 127–8)

There are at least two ways of denying that these are cases of punishment without accepting Hanna's view. One is derived from the account of punishment I outlined in the first paragraph of the paper; the other on considerations peculiar to the expressive view.

Both quarantine and involuntary confinement could take forms which would be found burdensome by a typical individual of the kind on whom it is being imposed. So for me, they would count as harsh treatment. However, it does not follow that they are forms of punishment.

On the account of punishment outlined in the first paragraph of the paper, a form of harsh treatment only constitutes punishment if it is imposed as a response to wrongdoing.¹⁰ If so, then quarantine is not (normally) a form of punishment, since it is not normally imposed as a response to wrongdoing.¹¹

In considering involuntary hospitalization of the mentally ill, we should distinguish between two different kinds of situation. In one an individual is confined on the grounds that they represent a danger to themselves or others. (In the UK, cases of this sort typically require invocation of Section 2 of the Mental Health Act and the signature of more than one doctor). Cases of this sort should not be regarded as cases of punishment because, like quarantine, they do not occur as a response to wrongdoing.¹²

The case of individuals who are sentenced to long-term confinement to secure psychiatric institutions after having committed a crime is more complicated. Such treatment is typically, in my terms, harsh; *ex hypothesi* it is a response to wrongdoing; and it is imposed by a appropriate authority. Unlike the cases of quarantine and of sectioning under the Mental Health Act, it is not immediately obvious that these should not be counted as cases of punishment. We should take seriously the possibility that these cases do involve punishment—maybe punishment masquerading as, or accompanied by something else. If they do, we may, rightly, be troubled by them.

Alternatively, one might argue that contrary to first appearances, these kinds of cases do not involve inflicting harsh treatment *in response* to wrongdoing. It is certainly true that in these kinds of cases, some apparent wrongdoing is the occasion of the imposition of harsh treatment. But there are good reasons for distinguishing between what occasions harsh

¹⁰ As I say in section VI, the word 'response' is normatively freighted here.

¹¹ It is also not, typically, imposed by the right kind of authority. Still, we can imagine cases where quarantine was judicially imposed which would still not be punishment.

¹² Hearings under the Mental Health Act, or similar legislation in other countries might be regarded as having some kind of quasi-judicial aspect to them. But, importantly, doctors are not asked to make any judgment as to whether there has been any kind of wrongdoing, but only of whether an individual is dangerous or not.

treatment and what the hard treatment in question is a response to. What occasions the harsh treatment is a breach of some social norm, which initiates a judicial and medical process, which results in confinement. However, it is arguable that any part of the harsh treatment which is not properly viewed as punishment is only legitimately imposed in cases where the individual is judged to be less than fully culpable because of their mental condition. In other words the imposition of non-punitive harsh treatment would typically be accompanied by a form of official recognition that this harsh treatment is precisely *not* a response to wrongdoing, but something else instead, such as a precaution against future danger.¹³

A second reason for thinking that standard cases of quarantine and confinement of the mentally ill do not constitute punishment can be extracted from views held by at least some of the expressivists against whom Hanna is arguing—and in particular by Feinberg and Duff. Duff and Feinberg both hold that even where harsh treatment is a response to wrongdoing, it only constitutes punishment if it either serves the right kind of expressive purpose or is intended to do so. (Feinberg 1970; Duff 2009)¹⁴ It is plausible that neither quarantine nor involuntary confinement of the mentally ill meet this condition. If they did meet this condition we would be rightly troubled by them, as we would in the case of involuntary confinement of the mentally ill imposed as a response to wrongdoing by non-judicial authorities.

6 Arrest, Self-Defense and Just War

Several other examples of behavior which one might take to involve inflicting harsh treatment as a response to wrongdoing, are also not plausibly viewed as forms of punishment.¹⁵ Here are four: the arrest and pre-trial detention of criminals; self defense; the defense of others against attacks on them; just warfare; and court ordered forcible restitution of property by offenders. None of them should be regarded as problematic by someone who thinks that expressive goals form part of the definition of punishment.¹⁶

Arrest and pre-trial detention of suspected criminals certainly involve treatment which I call harsh. However, these are arguably not responses to wrongdoing, but to reasonable suspicions of wrongdoing, since the innocent can (legitimately) be arrested. One might object that by the same token post-trial incarceration is not a response to wrongdoing either, since one can be found guilty and incarcerated, even if one is in fact innocent. However, the cases are disanalogous.

To see why, notice that the word ‘response’ is normatively freighted here, in a way that other similar words (such as the word ‘reaction’ on some readings) is not. If I am responding to X, I must be acting in a way that is appropriately sensitive to whether X has occurred—for

¹³ We can infer that it is not a response to wrongdoing from the fact that we would not judge the response mistaken if we found that the individual who is being confined is not culpable. culpability is a necessary condition for wrongdoing but not for breach of a social norm. (The significant difference here then is between culpable wrong doing and (mere, perhaps non-culpable) breach of a social norm, not between ‘response’ and ‘occasion’. I thank an anonymous referee for this journal for insisting on this clarification).

¹⁴ For more about what the right kind of expressive purpose might be, see section VIII. (As one referee insisted, we might think that involuntary detention of the mentally ill expresses concern for the safety of the community).

¹⁵ As a referee for this journal suggested.

¹⁶ An alternative response, which I do not explore here for reasons of space, would be to argue that while these may be case of punishment, they are cases of punishment which are not justified or justifiable in virtue of their status as acts of punishment.

example by being prepared to treat my actions as mistaken and to make due reparation for them if it can be shown that X has not occurred, and so on.¹⁷

This explains how the cases of pre-trial and post-trial incarceration differ. Someone who is found guilty, imprisoned, and then shown to be innocent is thereby shown to have been wronged.¹⁸ Someone who is found innocent after being arrested and detained need not be: if they have been detained on grounds of reasonable suspicion, they cannot claim to have been wrongfully arrested. So we have some grounds for saying that post-trial incarceration is a response to wrong-doing, whereas pre-trial detention is not: in cases where there has been no wrongdoing, we take it that we have, *for that very reason*, responded wrongly in the case of post-trial incarceration, whereas we do not automatically take this to be the case in situations where there has been pre-trial arrest and confinement, in situations where there are reasonable grounds to suspect wrongdoing but not actual wrongdoing.

It is harder to argue that self-defense is a response to something other than wrongdoing than it is to argue that arrest and pre-trial confinement are.¹⁹ We might want to say that in situations where I respond defensively to something which does not involve wrongdoing, I should take my response to be mistaken. This is not clear-cut: perhaps justifiable self-defense only requires a credible threat of wrongdoing, not actual wrongdoing. However, we should not depend upon this point.

The minimal account of punishment from section I provides one kind of ground for denying that self-defense is a form of punishment. On that account, punishment needs to be inflicted by an appropriate authority. Still, there might be cases of self-defense by figures such as policemen and judges, who one might take to be appropriate authorities. To deal with such cases, we might say that, in cases where they are defending themselves, policemen and judges are not acting in their capacity as appropriate authorities. However, this response requires an adequate account of what an appropriate authority is. There is some risk the account might turn out to be circular: it might end up saying, for example, that an appropriate authority is an which can punish justifiably.

The suggestion that punishment only counts as such if imposed by an appropriate authority also fails to explain why some cases of the use of force in the defense of others should not be counted as punishment. Some defensive uses of force appear to be imposed by appropriate authorities. A policeman may be authorized to use force to protect innocent by-standers; and a member of an army may be authorized to use force in response to wrongful invasion of a piece of territory. One might also appeal to traditional forms of Just War Theory for further examples of the use of force,

¹⁷ One referee objected thus: 'Isn't punishment possible in reaction to the mere suspicion of wrongdoing? When Israeli soldiers bulldoze the house of the merely suspected terrorist, they appear to be punishing him.' Whether this is indeed punishment depends on whether it is really a 'response' to wrongdoing in the normatively freighted sense I have drawn attention to. Conceivably, it might be: for example, if it was true that, when the suspected terrorist turned out to have been wrongly suspected, the soldiers—or the state which employed them—was willing to make appropriate reparation. But we might instead have here a case of retaliatory violence masquerading as punishment (and masquerading for precisely the purpose of trading on our sense of the justifiability of punishment). Something similar goes for another example suggested by the same referee: the imposition of 'punishments' in authoritarian regimes after show trials: here the idea that we have some kind of masquerade is built into the very description of the case.

¹⁸ In civilized jurisdictions this is publicly acknowledged.

¹⁹ The case of self-defense provides no support for a conception of punishment that involves an intention to harm. A self-defensive action may involve an intention to harm, without ceasing to be self-defense. (It may be that I can only defend myself effectively by inflicting pain on my attacker, for example).

imposed by a rightful authority, in response to wrongdoing, which should not be counted as forms of punishment.

Even if we think that arrest, pre-trial detention, some forms of self- and other-defense and justifiable war should be regarded as forms of harsh treatment which are responses to wrongdoings imposed by appropriate authorities, expressivists—or at least expressivists such as Duff and Feinberg—need not be committed to regarding them as forms of punishment. For they can argue that none of these forms of behavior are supposed to have the kinds of expressive function which are essential to punishment. This is clearest in the case of arrest and pre-trial detention. Since it has not been established, until a trial has occurred, that the arrested and detained individual has committed a wrong, the arrest and detention of a suspected wrong-doer cannot properly be supposed to express social condemnation of the wrong which has been committed.²⁰ Similar points might be made about the defense of self or others: we can only describe behavior in this way if it is intended to prevent harm to self or others. To the extent that it is intended to do more than this, we should regard it as something else. (In some cases, as when a looter in a flooded city is shot to ‘show what’s what,’ we may have a case of something which is, in fact, unauthorized, and hence illegitimate punishment).²¹

7 Court-Ordered Restitution

This leaves forced restitution, which could constitute harsh treatment on my account; is a response to a wrongdoing; and could be imposed by a court, which would constitute an appropriate authority. If forced restitution is a form of punishment, as some think, (Cholbi 2010)²² there is no objection here. However, legal systems which allow for the possibility of court ordered restitution, often treat it as a civil, rather than a criminal matter: judgments on claims for restitution may only require a judgment based on a ‘balance of probabilities’ standard of proof rather than proof beyond reasonable doubt. But perhaps a revisionary response is appropriate: since court-ordered restitution is a form of punishment, we should in fact require a criminal level of proof here.

What makes a particular standard of proof appropriate? If a court’s task is only to determine how the burdens brought about by a particular action should be distributed, a probabilistic judgment seems appropriate. However, if the court’s task is in part to expressing a moral judgment on behalf of society it is important that it be shown that the wrong being condemned has been committed. The converse is also plausible: we are entitled to take the fact that a case is decided on the basis of a balance of probabilities as a reason for thinking that the judgment is not intended to have a communicative purpose and thus that it

²⁰ So-called ‘perp walks’ are, therefore, problematic, since they seem to have the sort of expressive function which is constitutive of punishment, but they come not after an arrest, not a finding of guilt.

²¹ If it is part of our notion of punishment that it be imposed by an appropriate authority, can we make any sense of the idea of illegitimate punishment? Yes. I can have authority in a particular area, yet exercise it in ways that go beyond what is legitimate. (I may have authority to punish students in my class for plagiarism. But if I whip a plagiarist, what I do is illegitimate: I have exceeded my authority. Still, it is punishment). Notice that we need not concede that everything we can reasonably describe as a case of legitimate punishment is a case of punishment. Sometimes, illegitimate punishment is no more a case of something which is genuinely punishment, but illegitimate than a fake banknote is a particular species of banknote. (The ‘masquerades of punishment’ mentioned in note 17 are ‘fake’ in exactly the same way as a fake banknote is fake).

²² I thank Nathan Hanna for drawing my attention to this piece.

does not constitute a form of punishment on the expressive conception. We might then argue that some cases of court-ordered restitution—as that is currently understood—are not cases of punishment precisely because they are not intended to serve the expressive goals which punishment, properly so-called, is supposed to have.

Boonin has suggested that existing practices of civil restitution might be extended to cover cases where restitution needs to take non-monetary forms to deal with so-called ‘secondary harms’ to victims. (Boonin 2008 p 231–5) In many such cases, it may not be permissible or feasible for someone other than the wrongdoer to undertake restitution on the wrongdoer’s behalf.²³ This ‘non-substitutability’ feature suggests that we would not have a purely civil matter here. If so, this would be a reason for treating these cases as requiring proof ‘beyond reasonable doubt’, and of having an expressive purpose which makes them, *contra* Boonin, into punishment .

8 Have We Saved Expressivism?

I have argued that harsh treatment need not be intended to cause suffering. I now need to argue that punishment can be expressive in the ways expressivists require in virtue of being harsh (and not, for example, only in virtue of being intended to cause suffering).²⁴

On a denunciatory view, the purpose of the harsh treatment that punishment involves is for a society to communicate to its members that certain norms are in force and transgressions against them are viewed seriously. (Wringe 2006) A form of treatment need not be intended to cause suffering in order to have this role. A response to wrongdoing can be understood in the right way even if it does not actually cause suffering. It may be enough that it is the sort of thing which members of the society would typically take to be burdensome. If so, then, it need not be part of our intention in punishing someone that they be made to suffer. We need only intend that the punishment be harsh in the sense I have outlined. To see this, consider our responses to a citizen who, knowing that certain kinds of judicial response to wrongdoing were, in my terms ‘harsh’, concerned themselves with the question of whether particular, named individuals were actually being caused suffering. It is plausible that, absent special circumstances, we would find this troubling: it seems unduly voyeuristic.

Something similar is true on Bennett’s view. For Bennett, punishment involves the expression, on the part of society of certain ‘reactive attitudes’. (Bennett 2008 247–74). But, claims Bennett, the state has no legitimate interest in whether the individual has actually repented. (Bennett 2008 pp 196–7). Importantly, we can express reactive attitudes in conventional ways; and inflicting treatment which is, in my sense, harsh, can be seen as doing just this. There is no further requirement that the treatment be intended to cause suffering to the particular individual who is being punished.

On Duff’s view, matters are more complicated. Duff holds that punishment can only be justified if the sorts of treatment which punishment characteristically involved are aimed at making an offender experience remorse for their wrong-doing.²⁵ Since it is unpleasant to suffer remorse, one might think that if we inflict harsh treatment on individuals with the goal

²³ For example, Boonin suggests that a highly successful burglar might be required, to restore his victims previous sense of security by wearing a device which allowed the police to monitor his location. Clearly no-one else could do this on behalf of the burglar.

²⁴ I thank a referee for this journal for insisting that I make this explicit here.

²⁵ Duff sometimes writes in ways that suggest that this is part of the definition of punishment, and sometimes in ways which do not make it definitional but which do make it integral to the justification of punishment. For an example of the latter see Duff 2001 pp xiv–xv. For an example of the former see Duff 2009

of causing them to feel remorse we are intentionally causing suffering. If I do A in order to bring about B, then I have intentionally brought B about. So it might seem that the characterization of harshness offered here does not enable Duff's view to escape Hanna's criticism.

Perhaps this is simply a reason to prefer Bennett's version of the expressive view to Duff's. However, as I noted in section II, Hanna's critique of expressivism is based on the idea that even if punishment has expressive goals, it is possible to achieve these goals without engaging in treatment which is intended to cause suffering. (Hanna 2008 p 143) If one of the goals of punishment is to cause people to experience remorse, and if doing something in order to cause them to experience remorse is intentionally causing them suffering, then there is, in fact, no way of achieving this goal without intentionally causing them suffering. In short, on this line of argument, Hanna has no case against Duff.

Anticipating this point, Hanna suggests that a communicative theorist of Duff's stripe see the goal of punishment as being to cause the offender to acknowledge their wrongdoing, rather than make them feel remorse. The important point here for Hanna's purposes is that, as he emphasises, one can recognise that one has done wrong without suffering remorse. (Hanna 2008 p 144) Duff might not accept this reformulation of his views, on the grounds that a recognition that one has done wrong which does not involve some sort of suffering would not achieve the goals which remorse can achieve. (One might regard the suggested form of recognition as constituting a form of sociopathy where serious wrongdoing was concerned). However, even if Duff accepts the proposed reformulation, Hanna's critique still fails. If punishment involves treatment which is, in my terms, harsh and which is aimed not at causing remorse, but at bringing about a recognition of wrongdoing (with remorse as, presumably, an inessential, but perhaps frequent by-product), then it need not involve an intention to cause suffering after all.

9 Types of Offender: Some Problems

One might think my account of harsh treatment fails because it relies on the unacceptably vague idea of a typical individual of the same kind as the offender. I have said nothing about which of the ways in which individuals are similar to one another are relevant to deciding whether a given form of treatment is harsh in a given case. Should we consider—for example—the 'type' of an offender to consist of all individuals who are convicted of the same crime? Or all individuals accused of the same crime? Or all individuals convicted of a particular crime of a particular socio-economic class? Or all individuals in the society?

Although the account of harshness that I have given is indeed underspecified in this respect, this is not a significant problem. For the purpose of giving an account of harshness which is acceptable to the theorist of punishment it is important that there should be *at least one* way of explicating the notion of harshness on which a punishment's being harsh does not entail that it be imposed with the intention of causing suffering. This is true if there will *one* possible way of dividing offenders up into classes that have typical members. If there is more than one, then so much the better.

For practical purposes, matters are different: we may find that there is a difference between the kinds of punishments that we impose if we take as our standard for harshness what would be found burdensome by a typical member of the society doing the punishing, or what would be found burdensome by a typical individual convicted of a particular crime, or by a typical individual of a particular social class. (The problem only arises when there are noticeable differences between what typical members of these types would find burdensome.

There need not always be. For example, there may be no difference between what a typical member of the class of individuals convicted of an offence would find burdensome and what a typical member of the class of individuals accused of that offence would find burdensome). However it is not immediately obvious that philosophical considerations, as opposed to political, or practical ones ought to solve this problem.

The appeal to typical individuals in my characterization of harshness might be thought to risk the imposition of sentences which were disproportionately light or disproportionately severe in their impact on a particular individual. For an individual might be unusually resilient in the face of certain treatment or unusually prone to suffer from it. In other words they might be individuals who suffer either much more or much less than the typical individuals of their type. (This might matter simply on grounds of fairness; but an expressive theorist might also see it as something that could undermine the expressive goals punishment).

How far this is a problem on a practical level would depend on how we assign offenders to 'types'. So we might think our assignment of individuals to types should be done in such a way as to minimise the extent to which problems of this sort arise. Unfortunately, it seems to be built into the very structure of the account of harshness which I have offered that cases like this should be able to arise—that is, that there might be differences between what a particular individual finds burdensome and what an individual of their type might find burdensome. Even if we overlook this, it seems unlikely to be possible in practice to divide individuals into types that correspond to the exact degree to which they would find any particular kind of punishment burdensome.

Similar problems arise for any punitive regime which attempts to assign sentences on the basis of general guidelines. Since any reasonably realistic punitive regime would do so, my account of punishment faces no more problems than any other account would. Still, the objection might be sharpened by suggesting that, whereas we ought to regard this as an imperfection in our actual sentencing regimes, resulting from practical limitations on our knowledge of the circumstances of particular offenders, the account of punishment I have been offering requires it.

However it is not clear that a good account of punishment should yield a justification for perfectly tailored punishments. The fact that mine does not is a strength not a weakness. It reflects the fact in punishing me, the state should treat me impersonally. As Bennett has argued, there seems something worryingly illiberal about the idea that the state should be concerned with the state of my individual soul. One might regard this point as a problem for any version of the communicative view. But as we have seen, Bennett argues this is not correct. (Bennett 2006, 2008 pp 196–7) The communicative theorist can regard the state's imposition of punishments as providing a way for offenders to undergo something which can formally be regarded as being a form of penance. If so, what matters is not whether an individual actually suffers, but whether they undergo something which would typically be regarded as burdensome.²⁶

Advocates of communicative theories of punishment sometimes argue that their account are preferable to consequentialist accounts because consequentialist accounts treat punished individuals as a means to an end. One might think the account of punishment that I have

²⁶ What is true of the state may not be true of other kinds of individuals or institutions who have the right to punish individuals under certain circumstances, such as parents, schools or religious communities. This will strike some defenders of expressivism as a problem, insofar as they favor an account of what punishment is that applies both to state and non-state punishments. However, I do not think that allowing for a slight difference in what we take punishment to be is especially counter-intuitive here, particularly when weighed against the benefits of having an account of state punishment which escapes Hanna's objection.

offered is open to a similar worry. For the account seems to require us to treat individuals who are punished not *as* individuals, but as representatives of particular types, and thus perhaps, not as ends in themselves, but as mere means.

The Kantian prohibition on treating individuals as means to an end is best understood either as a prohibition on treating them in ways those could not consent to; or as a prohibition on treating them in a way which does not respect their rational nature. Neither understanding rules out the possibility of regarding punished offenders as representatives of a type. There is no reason why a penitent offender could not regard themselves as being appropriately treated as a member of a type—particularly if they understood the rationale for that treatment as being along the lines which a communicative theorist of punishment would suggest.

Treating an individual as a member of a type involves no disrespect for their rational nature. A rational being is one that is capable of recognising and responding in ways that are sensitive to, the force of rational considerations. But different types of rational beings can recognise and be sensitive to such considerations in distinctively different ways. Recognising that an individual is responsive to rational considerations in one way rather than another may well be a way of regarding them as a representative of a type. Seeing someone as a rational being of a particular kind is not inconsistent with—or even in tension with—seeing them as a rational being: rather it requires it.

10 Theory and Practise

I have claimed that there is a useful conceptual distinction to be made between punishment's being harsh, and its being intended to cause suffering to the individual on whom it is inflicted; and that Hanna's critique of expressive theories of punishment fails because it ignores this distinction. However, one might argue that although there is a conceptual distinction between punishment's being harsh and its being intended to cause suffering there is no practical difference: any institution which reliably inflicts harsh treatment will either do so by intentionally inflicting harm on those it treats harshly, or will do so in ways which involve someone intentionally inflicting harm on someone else.

It is plausible, given my characterisation of harsh treatment, that no institution could inflict harsh treatment on a regular basis without it being foreseeable that some individuals will be caused suffering. But we can distinguish between an outcome's being foreseen and its being intended. One way of doing so is to ask what an agent's response would be if the foreseen event did not occur. Suppose, then, that the miraculous does occur: wrongdoers are treated in a way which would typically cause suffering to individuals like them—but—on a one in a million chance—they do not suffer. If the inflicting of suffering on these individuals is intentional, we should regard the punishment as having failed. And if it is possible to cause these individuals suffering by making them undergo treatment which is idiosyncratically designed in such a way to cause them suffering, even if it would not do so to a typical individual like them, we should prefer to do this than inflict treatment which is—on the view of harshness defended here—harsh.

Whatever one might think of institutions which treated people in these ways, they would be very unlike our existing punitive institutions. So it does seem possible to distinguish, in this case, between suffering being inflicted in a way that is foreseen and suffering being inflicted in a way that is intentional. And this is what is required to meet the objection.

Someone might nonetheless refuse to accept a difference between treating an individual in a way that involves intentionally inflicting suffering on them and one which involves doing

so foreseeably. In defending the viability of the conception of harsh treatment outlined here, I have only claimed that the intended/foreseen distinction makes a moral difference in a very restricted range of cases. One might accept everything said in this section and still think that in many cases there is no significant difference between inflicting harm on someone intentionally and doing so foreseeably. However, anyone who agrees with Duff in seeing the legitimacy of punishment as resting on the nature of the relationship between punishing institutions and punished individuals, and in requiring that it be one that treats the punished as individuals is likely to find the distinction between intentionally treating someone in a particular way and doing so foreseeably to be morally significant. Refusing to accept the distinction in this context risks begging the question against many advocates of the communicative view.²⁷

References

- Bennett C (2006) State denunciation of crime. *J Moral Philos* 3:288–304
- Bennett C (2008) *The apology ritual*. Cambridge University Press, Cambridge
- Boonin D (2008) *The problem of punishment*. Cambridge University Press, Cambridge
- Cholbi M (2010) Compulsory victim restitution is punishment. *Publ Reas* 2:85–93
- Duff RA (1992) Alternatives to punishment—or alternative punishments. In: Cragg W (ed) *Retributivism and its critics*. Franz Steiner, Stuttgart
- Duff RA (2001) Punishment, communication and community. Cambridge University Press, Cambridge
- Duff RA (2009) Can We punish the perpetrators of mass atrocities. In: Brudholm T, Cushman T (eds) *The religious in response to mass atrocities*. Cambridge University Press, Cambridge
- Epictetus (1928) *Discourses* 2 vols tr W.A. Oldfather Cambridge Mass, Harvard University Press
- Feinberg J (1970) Justice and personal desert. In: *Doing and deserving*. Princeton University Press, Princeton
- Hampton J (1992) Correcting harms versus righting wrongs: the goal of retribution. *UCLA Law Rev* 39:1659–1702
- Hanna N (2008) Say what? A critique of expressive retributivism. *Law Philos* 27:123–150
- Hanna N (2009a) Liberalism and the general justifiability of punishment. *Philos Stud* 145:325–349
- Hanna N (2009b) The passions of punishment. *Pac Philos Q* 90:232–250
- Hart H (1959) Prolegomena to the principles of punishment. *Proc Aristot Soc* 60:1–26
- Metz T (2000) Censure theory and intuitions about punishment. *Law Philos* 19:491–512
- Wringe B (2006) Why punish war crimes: Victor’s justice and an expressive justification of punishment. *Law Philos* 25:159–191

²⁷ I thank Nathan Hanna and two referees for this journal for useful discussion.