

Why Punitive Intent Matters

Forthcoming in *Analysis*

Nathan Hanna

Many punishment theorists think that punishment is intentionally harmful and that this makes it especially hard to justify (e.g., Berman 2008: 266-7, 2012: 143, Boonin 2008: 6-17, Gardner 2008: xxxii, xli, Sayre-McCord 2001: 503-10, Tadros 2011: 1, 265, Zimmerman 2011: 3-10, 164).¹ But explanations for the latter intuition often say questionable things about the significance of the intent to harm. This makes some theorists doubt the intuition (e.g., Chiao 2013: 733-747, Enoch 2007: 73, Kolber 2012: 7, Ristroph 2008: 1391-94, Wringer 2019: 9-10). I'll argue that there's another way to explain the intuition – call it the target intuition.

Here's the plan. In section 1, I'll clarify what punishment theorists mean when they say that punishment is *intentionally* harmful. In section 2, I'll discuss the standard explanations for the target intuition, identify problems with them, and give a different explanation. In section 3, I'll defend my explanation. And in section 4, I'll discuss some implications and objections. I'll just assume that punishment is intentionally harmful, since that's the dominant view in the literature (cf. Boonin 2008: 13-14, n14; critics of the view include Kolber 2012, Ristroph 2008: 1399-1400, Sher 2009: 764, and Wringer 2013, 2019).

1. Clarifications

In this section I'll clarify what punishment theorists mean when they say that punishment is *intentionally* harmful (I won't clarify the meaning of *harmful* here, since others have done so, e.g., Boonin 2008: 6-12). And I'll explain why getting clear on this is important. Note that I'm concerned with what punishment theorists *mean*, not with the

¹ Note that my concern is with *moral* justification specifically.

nature of intention. For all I'll say, these theorists might be misusing the concept of intention. Nothing important to my discussion will turn on this.²

There are at least two ways to understand the claim that an agent intentionally harmed a subject by acting in some way (cf. Scanlon 2008: 10-11, Zimmerman 2011: 9-10). We can take it as a claim about what the agent knew: she knew that her act would harm the subject and acted anyway. Or we can take it as a claim about why she acted: harming the subject was one of her motivating reasons for acting.³ Punishment theorists who say that punishment is intentionally harmful mean this in the reasons-based sense (cf. Gardner 2008: xxx, n28). There are two pieces of evidence for this.

First, there's the language that they use to clarify their meaning. For example, they say that punishment is imposed 'because' it's harmful, that it's imposed 'in order to' harm, and that its harmfulness is 'essential' to it, not an 'incidental' or 'accidental' or merely 'foreseeable' 'side effect' of it (for discussion and references see Boonin 2008: 13 and Sayre-McCord 2001: 503-4). These are just less explicit ways of saying that punishment is intentionally harmful in the reasons-based sense.

Second, there's the standard argument that punishment theorists give for the intentions claim. Here's a paraphrase of the argument (cf. Berman 2012: 142-3, Boonin 2008: 12-15, Honderich 1969: 1, Ten 1987: 15, Wasserstrom 1982: 476, Zimmerman 2011: 9-10).

² Thanks to a referee for pressing me here.

³ Note that this doesn't mean that the agent inflicts harm for its own sake. She might inflict it because she thinks that it will have good effects (cf. Boonin 2008: 14, Chiao 2013: 733, Enoch 2007: 70, Sayre-McCord 2001: 504, Zimmerman 2011: 10).

It's possible to treat two people in comparably harmful ways for similar reasons and only punish one of them. For example, it's possible to confine two people because they're killers and only punish one of them. That's what we're typically doing when we imprison a culpable killer and commit a non-culpable mentally ill killer in a psychiatric institution. The intentions claim is necessary to distinguish punitive from non-punitive treatment in such cases. So it's true.

Such arguments are valid only if the intentions claim is meant in the reasons-based sense. This is because the forms of treatment being compared are all intentionally harmful in the knowledge-based sense. The fact that punishment theorists defend the intentions claim in this way shows that they're using the reasons-based sense.

Getting clear on this is important because it's tempting to think that punishing people is just a matter of treating them in harmful ways – subjecting them to so-called *hard treatment* – in response to their failings (Feinberg 1965: 397 rejects this view but suggests that it's common; he also seems to have popularized the term *hard treatment*). On this simple view, we punish people when we do things like fine or confine them for wrongdoing or rule breaking. If the simple view is true, justifying punishment is only a matter of justifying these ways of using hard treatment (among others, Wellman 2012 frames the task of justifying punishment in this way). But if the intentions claim is true, the simple view is false. We can use hard treatment in such ways without punishing because we can do it without intending harm. To illustrate, we can fine thieves just to compensate their victims or confine killers just to protect others from them (cf. Boonin 2008: 213-76, Gardner 2008: xxxi-xxxii, Sayre-McCord 2001: 508-9, Zimmerman 2011: 165-75). The target intuition says that this matters morally: other things equal, it's harder to justify

inflicting hard treatment with the intent to harm than without. That is, justifying punishment requires justifying both hard treatment and the intent to harm (cf. Berman 2012: 143).

Next, I'll present the standard explanations for the target intuition, raise some problems for them, and give another explanation.

2. Explanations – Part I

There are two standard explanations for the target intuition (advocates of one or both include Berman 2008: 267, 279-81, Boonin 2008: 15-16, 28-9, 234, Gardner 2008: xxxi-xxxiv, xl-xli, Sayre-McCord 2001: 507, Tadros 2011: 1, 139-66, and Zimmerman 2011: 159-65). Here's a paraphrase of the first explanation.

Intentions Can Affect Permissibility *Directly*: Other things equal, bad intentions can make an otherwise permissible act impermissible. The intent to harm is a presumptively very bad intention. It makes punishment especially hard to justify by threatening to make punishment impermissible in this way.

Advocates of this explanation point to cases where they think intentions affect permissibility directly. Here's such a case: a pilot fighting a just war bombs a key enemy weapons facility knowing that this will kill an innocent civilian (Boonin 2008: 16). Some theorists think that the pilot's intentions can affect whether she acts permissibly. They think that, other things equal, she acts permissibly if she just intends to destroy the facility and impermissibly if she intends to kill the civilian. Advocates of the first explanation say that the intent to harm can make punishment impermissible in the same way. For example, they worry that punitive confinement can be impermissible even in cases where comparably harmful non-punitive confinement would be permissible.

For reasons that I'll get to, some advocates of the target intuition doubt this explanation. So they've given another one. Here's a paraphrase of it.

Intentions Can Affect Permissibility *Indirectly*: Whether an act is permissible depends on its effects, which depend on its intentions. Other things equal, the intent to harm is likely to make a harmful act more harmful than it would otherwise be. More harmful acts are more likely to be impermissible. The intent to harm makes punishment especially hard to justify by threatening to make it impermissibly harmful.

Advocates of this explanation point to cases where they think intentions affect permissibility indirectly. Here's such a case: the authorities confine a dangerous person, thereby harming her. In one version of the case, they merely intend to protect others from her. In a second version, they intend to harm her (Boonin 2008: 234). Some theorists think that she'll likely be harmed more in the second version. These theorists say that the authorities' intent to harm her will probably lead them to do things like confine her in harsher conditions. These theorists worry that the intent to harm can affect punishment in the same way and that this might make punishment impermissibly harmful.

Both explanations have important problems. The first appeals to a controversial claim: that intentions can affect permissibility directly. There's a longstanding debate about whether intentions can do this. Many philosophers think not (e.g., Scanlon 2008 and those cited in Tadros 2011: 140 and Wellman 2012: 382-3; for opposing arguments see Tadros 2011: 139-66). They'd reject the first explanation outright. If possible, it would be better to explain the target intuition in a way that they can accept. More generally, it would be better to explain the intuition without taking sides in controversial debates like this.

The second explanation is too speculative. Consider: we can intend to inflict varying degrees of harm. Even if it's plausible to say that intending to inflict lots of harm is likely to make a harmful act more harmful than it would be if it wasn't intended to inflict harm, it's not obviously plausible to say this about intending to inflict a little harm. If the authorities in the above case confine the dangerous person and intend to harm her only a little, it's not obvious that they'll likely harm her more than they would if they didn't intend to harm her at all. So the problem raised by the second explanation isn't obviously a problem for punishment per se. It might just be a problem for punishment that's intended to be especially harmful.

I'm not saying that these problems are insurmountable, but I worry that they are. If the standard explanations were the only ways to explain the target intuition, then anyone who shares my worries should doubt the intuition. The standard explanations aren't the only options here, though. We can see this if we think about why so many advocates of the target intuition endorse these explanations: they do so because they assume that justifying punishment is only a matter of showing that it's permissible. Michael Zimmerman puts the assumption like this.

I am concerned with what it would take for punishment to be morally justified. ...To say that an act is morally justified is to say that it is not morally wrong, that to perform it is at least morally permissible and perhaps even morally required.

(Zimmerman 2011: 23, 25; cf. 159-65; cf. Berman 2008: 265-7, Boonin 2008: ix, 1-2, 28-9, Sayre-McCord 2001: 502-3, Tadros 2011: 1-4)

My explanation rejects this assumption (I'll say why shortly). Here's my explanation.

Intentions Matter Regardless of Permissibility: Justifying punishment requires showing that hard treatment is permissible and that inflicting it with the intent to harm is morally acceptable. The intent to harm makes punishment especially hard to justify because it's a presumptively very bad intention for an act even if the act is permissible.

My explanation avoids the problems with the standard explanations because it doesn't say that intentions can affect permissibility.

Next, I'll defend my explanation.

3. Explanations – Part II

Start with my claim that the intent to harm is a presumptively very bad intention for an act even if the act is permissible. To see why this is true, consider a case.

Carrier: Patient is a carrier of a very dangerous and highly infectious disease.

Governor quarantines her. This protects others from catching the disease from her and it harms her, e.g., by severely restricting her liberty. Governor does this with the intent to harm her.

There's a strong reason to quarantine Patient. So Governor might be acting permissibly. But even if he is, his intention is still very bad. He acts culpably and viciously by acting on it – viciously in the sense of manifesting vices like contempt, cruelty, and malice. His intention is bad in these ways because it's a deeply inappropriate attitude toward harm. It treats the fact that his act will harm Patient as a reason to perform the act when it's a reason not to (cf. Nagel 1986:181-2, Yaffe 2016: 165).⁴ These claims generalize. The intent to harm is

⁴ Note that which vices are present in such cases will depend on why this is a motivating reason. For example, if Governor intends harm for its own sake, that seems cruel and malicious. And if he intends harm because he thinks that his act will have good effects in virtue of its harmfulness – e.g., he thinks that the harm will get

almost always a deeply inappropriate attitude toward harm. This is because the fact that an act will harm someone is almost always a reason not to perform it, even if it's permissible. So part of my explanation is true: the intent to harm is a presumptively very bad intention for an act even if the act is permissible.

You might object that these claims don't support the target intuition because Governor could still be acting culpably and viciously even if he doesn't intend to harm Patient, e.g., if he's merely indifferent to her being harmed.⁵ This objection is mistaken. Other things equal, acting with the intent to harm seems to make Governor more culpable than mere indifference would. It also seems to manifest certain vices to a greater degree, e.g., it seems to express more contempt and it seems more cruel. And it seems like it can manifest distinct and especially bad vices like malice. Again, these claims generalize. Other things equal, intending harm seems worse than mere indifference to harm. So my claims do support the target intuition.

The other part of my explanation says that justifying punishment requires showing that inflicting hard treatment with the intent to harm is morally acceptable. My argument for this starts with the claim that some attitudes can be justified or unjustified. Pamela Hieronymi plausibly defends this claim as follows.

Certain attitudes are subject to direct rational criticism: one can be asked to give one's reasons for them. One can be asked why one believes, intends, resents, supposes, or imagines... Such requests for justification can, in principle, be satisfied.

(Hieronymi 2005: 454)

others to take more care to avoid infection – then that seems cruel to Patient and seems to express contempt for her (because it suggests that using her in this way is acceptable).

⁵ Thanks to a referee here.

There are no relevant differences between these attitudes and the intent to harm. So it can be justified or unjustified too.⁶ This has implications for punishment's justifiability: because punishment is intentionally harmful, its justifiability depends on the justifiability of the intent to harm. If the intent to harm someone is unjustified, so is punishing her.

You might object that this argument is invalid, that the justifiability of intentions entails nothing about the justifiability of acts. This objection fails. It overlooks the fact that some types of act are partly constituted by certain intentions and that acts of punishment are like this (Boonin 2008: 9 and Sverdlik 1996: 347 give examples of other such acts). Being intentionally harmful is just part of what it is for an act to be an act of punishment. Such acts can plausibly be said to be justified only when their constitutive intentions are, at least in one sense of *justified*. An act is justified in this sense only if it's *broadly morally acceptable*, e.g., not wrong or culpable or vicious. This is a perfectly legitimate sense of *justified* – and an important one, as I'll soon argue.

Next, I'll discuss some implications of my discussion and some more objections.

4. Implications and Objections

My discussion has at least two important implications. First, some attempts to justify punishment are consistent with the claim that punishment is deeply morally objectionable. Second, some arguments against punishment can be significantly strengthened. I'll elaborate.

Regarding the first implication, some attempts to justify punishment only argue that hard treatment is permissible (e.g., Wellman 2012: 371-6; Sher 2009: 764 gives other examples). Because of this, they don't suffice to justify punishment in my sense of *justify* –

⁶ Again, I mean *morally* justified, i.e., I'm concerned with moral reasons specifically.

remarkably, they're consistent with the claim that punishment is never morally acceptable and that punishers always act culpably and viciously. Anyone who wants to use these arguments to justify punishment must combine them with arguments that defend the intent to harm.⁷ In light of this, advocates of these attempted justifications – and punishment theorists more generally – should think more carefully about how if at all the intent to harm can be justified.

Regarding the second implication, some punishment theorists are abolitionists: they appeal to the target intuition and the standard explanations for it to argue that punishment is unjustified (e.g., Boonin 2008: 15-16, 28-9, 234, Zimmerman 2011: 159-65; cf. Sayre-McCord 2001: 507). Earlier, I outlined some objections to these explanations. Abolitionist arguments that appeal to the target intuition and rely on these explanations are vulnerable to the objections. The arguments can avoid the objections by appealing to my explanation instead. So we shouldn't be too quick to dismiss these arguments on the basis of objections to the standard explanations.

I'll conclude by considering two more objections to my arguments. The first says that my explanation for the target intuition doesn't actually support it because it might be false even if my explanation is true. To see the alleged problem here, recall the intuition: the intent to harm makes punishment especially hard to justify. And recall my explanation, in brief: justifying punishment requires justifying both hard treatment and the intent to harm. This explanation doesn't rule out the possibility that these things can be justified in

⁷ This concern might apply to many attempts to justify punishment. Boonin argues that many attempts don't show that acting with the intent to harm is permissible (Boonin 2008: 35, 117-19, 141-43, 147-49, 170-71, 205-207). His arguments can be reframed in terms of acceptability. See below.

the same way. If they can be, the objection says, then the intent to harm won't make punishment especially hard to justify.⁸

This objection fails because the previous sentence is false. To see why, consider the reasons that might justify hard treatment and the intent to harm. There seem to be a variety of reasons that can justify hard treatment.⁹ But cases like *Carrier* seem to show that some of these reasons can't justify the intent to harm at all or can't justify it as often as they can justify hard treatment.¹⁰ This apparent fact about the intent to harm can explain why it makes punishment especially hard to justify even if it can be justified in the same way as hard treatment. So the first objection fails.

The second objection says that it doesn't much matter – that there's not much reason to care – whether punishment is justified in my sense, so long as we can permissibly use hard treatment to punish. You might even say that *Carrier* suggests as much, that Governor's bad intention doesn't much matter given the obvious reason to think that he acts permissibly: the fact that his act is necessary to avert a grave threat.¹¹

Two points in reply. First, the objection is wrong about *Carrier*. The fact that Governor's act is necessary to avert a grave threat doesn't show that his bad intention doesn't much matter. At most, this fact just shows that something about his act matters more than his bad intention and that performing the act with the bad intention is better

⁸ Thanks to a referee for this objection.

⁹ For example, if hard treatment has very good consequences – e.g., that outweigh a persons' rights or interests – that can help to make it permissible. And if it doesn't violate a person's rights – e.g., if he waives or forfeits rights against it – that can help to make it permissible (cf. Wellman 2012).

¹⁰ To illustrate, the hard treatment in *Carrier* seems justified partly because it has very good consequences. And maybe it's justified partly because it doesn't violate Patient's rights. But these considerations don't justify Governor's intention. At the least, this suggests that these considerations can't justify the intent to harm as often as they can justify hard treatment. I suspect that good consequences can justify the intent to harm in certain special cases and that facts about rights never can, but I won't explore these issues here.

¹¹ Thanks to two referees for this worry about *Carrier*.

than not performing it at all. This is consistent with the claim that his bad intention still matters in important ways and that performing the act without the bad intention would have even better.

Second, there are good reasons to think that Governor's intention – and the intent to harm more generally – do matter in important ways. Some remarks of P.F. Strawson's help to show why.

The central commonplace that I want to insist on is the very great importance that we attach to the attitudes and intentions towards us of other human beings ... [I want to] emphasize how much we actually mind, how much it matters to us, whether the actions of other people ... reflect attitudes towards us of goodwill, affection, or esteem on the one hand or contempt, indifference, or malevolence on the other. (Strawson 2008: 5-6; cf. Zimmerman 2011: 162-4)

Strawson's point is just that we care a lot about people's attitudes and intentions. But I would add that we have good reasons to care. To see this, consider Carrier again. Even if Governor acts permissibly, his bad intention still matters in several ways. Here are some. It's deeply disrespectful to Patient. It justifies attitudes towards Governor like resentment, indignation, and blame – as well as guilt and shame on his part. And it can justify acts that express these attitudes, e.g., criticizing him and voting against him – as well as apologizing and resigning on his part. These claims generalize. Even when hard treatment is permissible, whether it's unjustifiably intended to harm still matters in important ways. So the second objection fails too.¹²

¹² I've focused on persons who intend harm, but there are arguably cases where collectives like states intend harm. You might wonder if my points apply in such cases. For reasons to think that they do see Tollefsen (2003). Thanks to a referee for raising this issue and directing me to Tollefsen's paper.

Conclusion

The target intuition says that the intent to harm makes punishment especially hard to justify. The standard explanations for this intuition are problematic. So I've defended another one.¹³

References

- Berman, M. 2008. Punishment and Justification, *Ethics*, 118: 258-290.
- Berman, M. 2012. The Justification of Punishment, in *The Routledge Companion to Philosophy of Law*, ed. A. Marmor, New York: Routledge.
- Boonin, D. 2008. *The Problem of Punishment*, New York: Cambridge University Press.
- Chiao, V. 2013. Punishment and Permissibility in the Criminal Law, *Law and Philosophy*, 32: 729-65.
- Enoch, D. 2007. Intending, Foreseeing, and the State, *Legal Theory*, 13: 69-99.
- Feinberg, J. 1965. The Expressive Function of Punishment, *The Monist*, 49: 397-423.
- Gardner, J. 2008. Introduction, in H.L.A. Hart, *Punishment and Responsibility*, 2nd ed., New York: Oxford University Press.
- Hieronymi, P. 2005. The Wrong Kind of Reason, *Journal of Philosophy*, 102: 437-57.
- Honderich, T. 1969. *Punishment: The Supposed Justifications*, New York: Harcourt, Brace & World.
- Kolber, A. 2012. Unintentional Punishment, *Legal Theory*, 18: 1-29.
- Nagel, T. 1986. *The View from Nowhere*, New York: Oxford University Press.
- Ristroph, A. 2008. State Intentions and the Law of Punishment, *Journal of Criminal Law and Criminology*, 98: 1353-1406.

¹³ Thanks to Craig Agule, Robert Hughes, Gerard Vong, and several anonymous referees for helpful comments. Thanks also to audiences and workshop participants at CU Boulder, Drexel University, Northwestern University Law School, Roskilde University, the 2020 Eastern APA, the 2016 and 2018 Northwest Philosophy Conferences, the 2016 Pittsburgh Area Philosophy Colloquium, and the 2017 Rocky Mountain Ethics Congress.

- Sayre-McCord, G. 2001. "Criminal Justice and Legal Reparations as an Alternative to Punishment," *Philosophical Issues*, 11: 502-29.
- Scanlon, T.M. 2008. *Moral Dimensions*, Cambridge, MA: Harvard University Press.
- Sher, G. 2009. *The Problem of Punishment*, *Ethics*, 119: 761-764.
- Strawson, P.F. 2008. Freedom and Resentment in P.F. Strawson, *Freedom and Resentment and Other Essays*, New York: Routledge.
- Sverdlik, S. 1996. "Motive and Rightness," *Ethics*, 106: 327-49.
- Tadros, V. 2011. *The Ends of Harm*, Oxford: Oxford University Press.
- Ten, C.L. 1987. *Crime, Guilt, and Punishment*, Oxford: Clarendon Press.
- Tollefsen, D.P. 2003. Participant Reactive Attitudes and Collective Responsibility, *Philosophical Explorations*, 6: 218-34.
- Wasserstrom, R. 1982. Capital Punishment as Punishment: Some Theoretical Issues and Objections, *Midwest Studies in Philosophy*, 7: 473-502.
- Wellman, C.H. 2012. The Rights Forfeiture Theory of Punishment, *Ethics*, 122: 371-93.
- Wringe, B. 2013. Must Punishment Be Intended to Cause Suffering?, *Ethical Theory and Moral Practice*, 16: 863-77.
- Wringe, B. 2019. Punishment, Judges, and Jesters: A Response to Nathan Hanna, *Ethical Theory and Moral Practice*, 22: 3-12.
- Yaffe, G. 2016. Desert for Wrongdoing, *Journal of Ethics*, 20: 149-71.
- Zimmerman, M. 2011. *The Immorality of Punishment*, Peterborough, Ontario: Broadview Press.