Ethic Theory Moral Prac
DOI 10.1007/s10677-017-9835-9

The Nature of Punishment: Reply to Wringe

Nathan Hanna[1]

Accepted: 12 September 2017

1 The Nature of Punishment: Reply to Wringe

David Boonin and I have independently argued that many justifications of legal punishment fail. Our arguments rely partly on the claim that an agent punishes a subject only if the agent aims to harm the subject. Call this the *Aim to Harm Requirement* or *AHR* for short. Boonin says that it's "almost universally accepted" in the punishment literature (Boonin 2008: 13-14, n14; cf. Hanna 2008: 126, 2009: 330: 593). And we've both argued for it at length even though few philosophers object to it in print (Boonin 2008: 6-21; Hanna 2008: 125-8, 2009: 329-31). Here, I'll consider the arguments of one of the few who has.

Wringe (2013) defends his preferred justification of legal punishment against our criticisms by arguing that AHR is false.[1,2] His arguments are original. And like-minded philosophers have endorsed them (e.g., Glasgow 2015: 612, n28; Lee 2017). In this paper, I'll show that Wringe's arguments against AHR fail. By my count he gives two such arguments. The first attacks AHR directly. The second attacks certain arguments for it.

2 Argument One: Against AHR

Wringe's first argument cites a popular account of legal punishment and tries to show that a plausible reading of it is consistent with denying AHR. He credits the account to H.L.A. Hart and puts it like this: legal punishment "involv[es] harsh treatment [and is] inflicted on an offender by an appropriate authority, in response to … wrongdoing" (863, citing Hart 1968: 4-5). Wringe takes the reference to harsh treatment here to be especially important and tries to clarify what it

[1]All references are to Wringe (2013) unless otherwise noted. Wringe (2016: 18-41) reiterates the arguments.

[2]Another objection to our arguments grants AHR and says that the aim to harm doesn't have the moral significance that we take it to have. We both address this objection (Boonin 2008: 15-16, 28-29, 61-2, 234; Hanna 2008: 333, 2014: 595-7). Wringe seems to press a version of it at one point, but he doesn't consider our replies (868-9). I give a novel reply in Hanna (ms).

✉ Nathan Hanna
nhanna@drexel.edu

[1] Drexel University, Philadelphia, PA, USA

Published online: 20 September 2017

means. An obvious interpretation takes harsh treatment to be harmful treatment. But Wringe says that there's another interpretation: on his view, harsh treatment is "treatment which would normally be found burdensome by a typical individual of the kind on whom it is being imposed" (867).[3] He notes that this characterization has an interesting implication: because treatment that's only normally harmful can fail to harm, we can treat someone harshly without harming her or aiming to. In light of this, Wringe concludes that a plausible reading of Hart's account is consistent with the claim that we can punish someone without harming her or aiming to. On this reading, an alternative to AHR is true: an agent doesn't have to aim to harm a subject to punish the subject; the agent just has to aim to treat the subject harshly in Wringe's sense (868).[4]

There are two problems with this argument. First, Wringe misstates Hart's account. Hart doesn't say that legal punishment involves *harsh treatment*. What he says is that legal punishment "must involve pain or other consequences normally considered unpleasant" (Hart 1968: 4). Later, he just says that legal punishment is "painful" and that it "entails suffering" (ibid.: 6, 26).[5] These claims are inconsistent with the claim that legal punishment can be harmless.[6] Since Wringe misstates Hart's account, he hasn't shown that a plausible reading of it is consistent with denying AHR. That said, his take on Hart's account might be independently plausible even if it's not what Hart had in mind. This brings me to the second problem with Wringe's argument: there are good reasons to reject his take on Hart's account.

Recall, Wringe's take on Hart's account says that legal punishment "involv[es] harsh treatment [and is] inflicted on an offender by an appropriate authority, in response to … wrongdoing." Wringe says that harsh treatment can sometimes be harmless because it only has to be normally harmful. And his alternative to AHR denies AHR and says that an agent must aim to treat a subject harshly in order to punish. This account has false implications. Consider:

> **Judgment**: Judge must sentence Thief, but doesn't want to harm him. Giving him an obviously harmless sentence will anger the public, though. To avoid this, she gives him a sentence that would harm the vast majority of people but that – for reasons known only

[3] Wringe shifts between talk of burdensome treatment, harmful treatment, and treatment that causes suffering. He also shifts between talk of aims and intentions. He uses these terms interchangeably. I'll put my claims in terms of the aim to harm. I'll preserve Wringe's wording when quoting him and use mine when paraphrasing.

[4] Much of the rest of his paper defends three claims. One: the distinction between aiming to harm a subject and aiming to treat the subject harshly is a genuine distinction (867-8). Two: a harshness based account of legal punishment can accommodate our intuitions about cases (869-73). Three: his preferred "denunciatory" account of legal punishment is consistent with his alternative to AHR (873-4). I'll focus on the second claim. I'll argue that there are counterexamples to harshness based accounts and that they don't accommodate important intuitions about cases.

[5] As I read Hart, he's saying that punishment must be bad for its victim (harmful) and that being painful is the most obvious way it can be so. He credits the account to Benn and Flew, who are more explicit about this. Flew says that punishment must be an "evil, an unpleasantness, to the victim" and that it needn't be painful (1954: 293). Benn says that it must "involve" such an evil (1958: 325). More recently, Boonin gives good reasons to characterize punishment in terms of harm rather than pain and suffering (2008: 6-7). That said, pain and suffering are salient harms that punishment is often meant to inflict. So it's understandable why Hart characterizes punishment in these terms (cf. Hanna 2008: 126).

[6] Hart says that punishment *involves* pain or unpleasantness. A referee suggested that this is consistent with Wringe's alternative to AHR because the alternative captures a way that punishment can be said to *involve* such things. I don't think that this is a plausible reading of Hart. It's inconsistent with his later, more straightforward characterizations of punishment. And it takes his criterion to be stating a merely typical feature of legal punishment rather than a necessary one. Hart seems to intend his criterion in the latter way (his ensuing discussion of non-legal punishments suggests this). I think that his use of the word *involves* is just meant to leave open the nature of the relationship between this criterion and his other criteria. There's no indication that it's meant to leave open the possibility that punishment can be harmless.

> to the two of them – won't harm him. This will avoid public anger because people will just assume that the sentence harms him.

This sentence satisfies the above criteria. It's imposed on an offender by an appropriate authority in response to wrongdoing. It's harsh in Wringe's sense, since it would harm the vast majority of people. And the authority aims to treat the offender harshly in Wringe's sense. So the account entails that the sentence is a punishment.[7] But it's not. It's just meant to look like a punishment. So the account is false. And AHR plausibly explains where the account goes wrong: Judge doesn't punish Thief because she doesn't aim to harm him.

Wringe might object that Judgment is too abstract and that a concrete version of it wouldn't obviously involve a harmless sentence. The best way to deal with this worry is to make the case more concrete. But the best way to do that depends on the right theory of harm. Since I can't defend such a theory here, I'll just give a more concrete version of the case in line with what I take to be a plausible theory.[8] Suppose then that Thief loves doing community service. He doesn't find it at all unpleasant. And if he's able, he'll do lots of it after he's sentenced. Judge knows all of this and sentences him to do the very kind of community service that he was planning to do. Thief eagerly does even more than he's required to do, greatly enjoys it, and doesn't find it at all unpleasant. To me at least, it's obvious that Thief isn't harmed and that Judge doesn't punish him. I suspect that many would agree.[9]

There's another way that Wringe might try to resist my counterexample, though. He might try to further clarify the nature of harshness in an attempt to show that Judge doesn't really treat Thief harshly (cf. 874-6). If she doesn't, then Wringe's take on Hart's account won't entail that she punishes Thief. Recall, Wringe says that harsh treatment is "treatment which would normally be found burdensome by a typical individual of the kind on whom it is being imposed." Here are two representative versions of the reply.

- Whether Judge treats Thief harshly depends on her psychology. She aims not to harm him and imposes a sentence that she knows won't harm him. Treating people like *that* is never harmful. So it's not normally harmful. So it's not harsh.
- Whether Judge treats Thief harshly depends on facts about Thief. A typical individual like him would have the personal characteristics that make the sentence harmless to him. So treating him in this way isn't normally harmful. So it's not harsh.

These replies characterize harsh treatment in ways that are designed to evade my counterexample. But they don't explain why their characterizations are good ones. So they're ad hoc.

[7] Wringe's preferred denunciatory account of legal punishment says that an act is a legal punishment only if it's meant to publicly denounce wrongdoing. I'll discuss this claim later. For now, just notice that it poses no problem here. We can stipulate that Judge imposes the sentence to (insincerely?) denounce Thief's thievery in a non-harmful way. And we can stipulate that being denounced in this way doesn't harm Thief, e.g., because he finds the attention paid to his daring thefts enjoyable and in no way unpleasant.

[8] I defend my preferred theory in Hanna (2016). It's not clear what theory Wringe accepts. But he seems to endorse some odd claims about harm. He suggests that a subject who genuinely believes that imprisonment can't harm her wouldn't be harmed by being imprisoned (867). No mainstream theory of harm makes harm belief-sensitive in this way. For a critical overview of theories see Bradley (2012).

[9] A referee objected that the sentence might still be a punishment because being sentenced to do community service has a certain "social meaning." In reply, Judgment seems to show that a token sentence can share the social meaning that sentences of its type typically have without being a punishment. To put this point in terms of the particular meaning that Wringe is concerned with, a token sentence can serve to publicly denounce an offender's conduct without being a punishment. Judge's sentence may have such a meaning only because the public is mistaken about the details of the case.

If Wringe wants to resist my counterexample by clarifying the nature of harshness, he must give a principled account of harshness that can figure in a plausible account of punishment. But there's a good reason to think that this can't be done: there doesn't seem to be a necessary connection between how a form of treatment would normally affect people and whether a particular instance of it is a punishment. Consider: it seems possible to determine whether an agent's treatment of a subject is a punishment just by considering facts about the agent and the subject such as the agent's motives and the effects of the treatment on the subject. And it seems that the treatment can be a punishment even if the subject is unique and no other subject would be harmed by being treated so treated. Wringe doesn't give any good reasons to deny these claims. So his alternative to AHR seems unmotivated.

You might disagree. You might think that there are good motivations for Wringe's alternative. Two facts in particular might look like good reasons to accept it. The first is that state officials typically punish offenders by using standard sentences that would harm most people. This isn't a good reason to accept Wringe's alternative, though because AHR is consistent with this fact and can help to explain it. State officials do this because it's an efficient and sufficiently reliable way to harm offenders, given limited time and resources.

Another fact that might look like a good reason to accept Wringe's alternative is this: his alternative entails that the arguments that Boonin and I give against legal punishment fail. This is because our arguments rely on AHR. Wringe himself says that this is a good reason to accept his alternative even if it has counterintuitive implications. Specifically, he says that the claim that we can legally punish without harming or aiming to harm might seem counterintuitive, especially to those of us who find an analogous claim about non-legal punishment counterintuitive. But he insists that accepting these implications is worth being able to reject the arguments that Boonin and I give. Here's how Wringe puts this point.

> What is true of the state [on Wringe's alternative] might not be true of other kinds of individuals or institutions … such as parents, schools, or religious communities. This will strike some … as a problem, insofar as they favor an account of what punishment is that applies to both 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 [and Boonin's] objections. (875, n26)

The purported benefit here isn't a good reason to accept Wringe's alternative. Treating it that way violates a plausible criterion for a good account of legal punishment: that the account be *neutral* about whether and why legal punishment is morally justified.[10] Wringe doesn't give any good reason to reject this criterion. So this motivation for his alternative is a bad one.

I conclude that Wringe's first argument against AHR fails. His harshness based take on Hart's account is false and his alternative to AHR is unmotivated. I'll discuss his second argument in a moment. But first I want to briefly discuss two objections to AHR that he doesn't make.[11] The first says that AHR is false because institutions of punishment can be structured so that officials can punish without aiming to harm. To illustrate, suppose that a judge is legally required to impose a mandatory minimum prison sentence on a convicted defendant and that she does so only because she's legally required to, not because she aims to harm him. AHR entails that she doesn't punish

[10] See Bedau and Kelly (2015: Sec. 2), Boonin (2008: 5-6), and Zimmerman (2011: 1-2).
[11] Thanks to two referees for raising these objections.

him, but you might think that she does. The second objection says that AHR has false implications in a variant of Judgment. Suppose that Judge's sentence harms Thief despite her aim not to harm him. AHR entails that she doesn't punish him. But you might think that she does.

These are challenging objections. They show that AHR's advocates have to think more carefully about the nature of punishment in complex institutional contexts. I'm optimistic that the objections can be overcome, but I won't try to do that here. I'm just trying to show that Wringe's attack on AHR fails. Since he doesn't raise these objections, I won't try to defend AHR against them.[12]

This might seem evasive, though. You might think that Wringe's alternative to AHR avoids such objections and that this gives it a clear advantage over AHR.[13] Not so. Wringe's alternative is vulnerable to similar objections. Consider: a judge can inadvertently impose a harsh sentence while aiming not to. And it's possible for a judge to impose a legally mandated sentence without aiming to treat an offender harshly. You might think that judges who act in these ways are still punishing. But Wringe's alternative entails that they're not. The failure to realize this stems from a misunderstanding of his alternative. Recall, his alternative says that an agent punishes a subject only if the agent aims to treat the subject harshly. Wringe explicitly says that an agent doesn't do this just by aiming to treat the subject in a way that happens to be harsh. To satisfy the criterion, Wringe says, the agent must treat the subject in that way *because it's harsh* (868). In other words, the harshness of the treatment must be among the agent's reasons for imposing it.[14] So Wringe's alternative is no better off here.

Now I'll consider Wringe's second argument against AHR.

3 Argument Two: Against Arguments for AHR

Wringe's second argument against AHR attacks one of my arguments for it. Here's how Wringe puts my argument.

> 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 [punishment must involve] 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. (869, citing Hanna 2008: 127-8)

Wringe says that there are two ways to reject this argument. One: say that some treatment is a legal punishment only if it's a response to wrongdoing (869-70). Two: say that some treatment is a legal punishment only if it's meant to denounce wrongdoing in a sense spelled out by his "denunciatory" account of legal punishment (870, n14). That account says that "the purpose of the harsh treatment that [legal] punishment involves is for a society to communicate to its members that certain [moral] norms are in force and that transgressions against them are viewed seriously" (865, 873). Things like quarantine and involuntary psychiatric commitment

[12] Michael Zimmerman offers one plausible reply to such worries (2011: 19-21). He says that AHR is true only for agents who act on their own behalf and that a variant of it is true for agents who act on behalf of others like the state. According to this variant, an agent who doesn't aim to harm a subject can still punish the subject by acting on behalf of someone else who aims to harm the subject. I'm inclined to take a harder line against the above objections and say that AHR s true for everyone. But I won't give my preferred reply here.

[13] Thanks to a referee for raising this worry.

[14] I frame AHR in such terms in Hanna (ms).

typically don't satisfy these criteria. So the criteria entail that they're typically not legal punishments.

There are two problems with this argument. First, Wringe misstates the argument that he's citing. In the passage that he cites, he takes me to be arguing that AHR is true because it counts the above forms of treatment as non-punitive. But that's not what I says in the cited passage. Here's what I say.

> The claim that the aim to impose suffering is essential to punishment has a certain intuitive appeal, but there are other considerations that speak in its favor. The aim invariably influences the way that punishments are applied and so helps to account for significant differences between punitive and non-punitive treatment, e.g., differences between [punitive] imprisonment and other kinds of confinement. Medical quarantine, involuntary psychiatric commitment and protective confinement all cause suffering. Ideally, however, steps are taken to minimize the suffering they cause. This is not the case with [punitive] imprisonment. Offenders are imprisoned, at least in part, in order to make them suffer and prison conditions are designed in service to this aim. (Hanna 2008: 126; Hanna says that by *suffering* he means *hurt or harm*)

In this passage, I argue for AHR by appealing to its explanatory value: it explains some typical differences between some punitive and non-punitive forms of confinement. Wringe's wrongdoing and denunciation criteria don't do this. The former doesn't because it doesn't say *how* legal punishment responds to wrongdoing. The latter doesn't because it doesn't explain *how* legal punishment denounces wrongdoing. So neither criterion poses a problem for what I actually say in the above passage. Of course, the *conjunction* of these criteria with Wringe's alternative to AHR can in principle do some of this explanatory work. But we've seen that there are good reasons to doubt Wringe's alternative. So appealing to it isn't a good way to do that work.

That said, the explanatory argument isn't the only one that I give. Elsewhere, I do argue for AHR by saying that it correctly classifies certain forms of treatment as non-punitive. But a crucial part of that argument says that accounts of punishment that try to do this while denying claims like AHR have false implications (Hanna 2009: 330-31). The second problem with Wringe's argument is that the wrongness and denunciation criteria have such implications. Consider:

> **Bombed**: Comic is performing a stand-up comedy routine for King, a mercurial monarch with absolute legal authority. King's unpredictability makes her nervous. As a result, her timing is off and her jokes aren't funny. King is displeased by this. He sentences her to a week in the royal dungeon for not being funny. But he doesn't think that she has done anything morally wrong and isn't morally denouncing her actions.

King punishes Comic. There's nothing odd about this. Punishing someone without thinking that her actions were wrong and without morally denouncing her actions is surely possible. To take just two examples, a dog owner can do this to her dog for peeing inside and a piano teacher can do this to a new student for a slip of the fingers – say with a rap on the nose and the knuckles, respectively. Moreover, King *legally* punishes Comic because he exercises his legal authority in doing so. Wringe's criteria entail that King doesn't legally punish Comice. So Wringe's criteria are false.[15]

[15] I'd say the same about a weaker version of the wrongdoing criterion, e.g.: an act is a legal punishment only if it's a response to a legal offense. Bombed is also a counterexample to this claim because Comic doesn't commit a legal offense (cf. Zimmerman 2011: 14). That said, a weaker version of the disapproval criterion might be true. Maybe all punishments express disapproval. Even if such a criterion is true, though, it doesn't seem helpful to Wringe. Such a criterion can't obviously do all of the explanatory and conceptual work that AHR does.

It's worth noting that other mainstream accounts of legal punishment eschew these criteria, presumably for similar reasons. To take a few examples, Boonin's account says that legal punishment is a response to a legal offense and that it expresses disapproval (2008: 6-26). He puts neither condition in moral terms. Hart's account also includes a legal offense condition, but has no disapproval condition (Hart 1968: 4-5; cf. Benn 1958: 325, Flew: 1954: 293). Even Feinberg's account – a forerunner of Wringe's denunciatory account – doesn't obviously require that the disapproval be moral. Feinberg says:

> It is much easier to show that punishment has a symbolic significance than to say exactly what it is that punishment expresses. *At its best*, in civilized and democratic societies, punishment … expresses the judgment … of the community that what the criminal did was wrong." (1965: 402-3, emphasis added)

Given all this, I see no reason to accept the wrongdoing and denunciation criteria. And I see no reason to think that appealing to them instead of AHR is a good way to do the relevant explanatory and conceptual work.

Wringe might reply that I've misunderstood the wrongdoing and denunciation criteria. He might say that they're not meant as constraints on what legal punishment is, but as constraints on what *morally justified* legal punishment is and that I've given no good reasons to reject them, so understood.[16] This reply might seem plausible, especially since Hart says something similar about his own account (Hart 1968: 5-6).[17] But the reply fails. So understood, Wringe's criteria are consistent with AHR and with many arguments for it. This is because AHR is a view about what punishment is, not a view about what morally justified punishment is. If we understand Wringe's criteria in the proposed way, they're just irrelevant to the present debate about AHR.

Given these problems, I conclude that Wringe's second argument against AHR also fails. Neither of his arguments give us any good reasons to doubt AHR.[18]

References

Bedau, H.A., and Kelly, E. (2015). "Punishment," The Stanford Encyclopedia of Philosophy (Fall 2015 Edition), Edward N. Zalta (ed.), URL = <https://plato.stanford.edu/archives/fall2015/entries/punishment/>
Benn SI (1958) An approach to the problems of punishment. Philosophy 33:325–341
Boonin D (2008) The problem of punishment. Cambridge University Press, New York
Bradley B (2012) Doing away with harm. Philos Phenomenol Res 85:390–412
Feinberg J (1965) the expressive function of punishment. Monist 49:397–423
Flew A (1954) The justification of punishment. Philosophy 29:291–307
Glasgow J (2015) The expressivist theory of punishment defended. Law Philos 34:601–631
Hanna N (2008) Say what? A critique of expressive retributivism. Law Philos 27:123–150
Hanna N (2009) Liberalism and the general justifiability of punishment. Philos Stud 145:325–349
Hanna N (2014) Facing the consequences. Crim Law Philos 8:589–604
Hanna N (2016) Harm: omission, preemption, freedom. Philos Phenomenol Res 93:251–273

[16] Some of his remarks suggest that he'd find this reply attractive. For example, he seems at one point to say that expressing certain messages is important because it's central to the justification of punishment, not necessarily because it's constitutive of punishment (865-6). This thought seems to influence later parts of his paper, e.g., his puzzling remarks about harm and voyeurism (873) and his remarks about justifying what he calls "perfectly tailored punishments" (875-6).

[17] At times, Feinberg seems to see his own account this way too (1965: 408-18). And he emphasizes at the start of his paper that he's using "a narrower, more emphatic sense" of *punishment* (1965: 397-8).

[18] Thanks to Stephen Galoob, Bill Wringe, and two anonymous referees for this journal.

Hanna N (ms). Punishment, Permissibility, and Justification. Drexel University
Hart HLA (1968). Prolegomenon to the principles of punishment. In: Hart HLA (ed) Punishment and responsibility. Oxford University Press, New York
Lee AYK (2017) An expressive theory of punishment. Philos Q 67:418–421
Wringe B (2013) Must punishment be intended to cause suffering? Ethical Theory Moral Pract 16:863–877
Wringe B (2016) An expressive theory of punishment. Palgrave MacMillan, Basingstoke
Zimmerman MJ (2011) The immorality of punishment. Broadview Press, Peterborough