Ethical Theory and Moral Practice
https://doi.org/10.1007/s10677-019-10047-1

The Nature of Punishment Revisited: Reply to Wringe

Nathan Hanna[1]

Accepted: 25 November 2019/Published online: 09 December 2019

Abstract
This paper continues a debate about the following claim: an agent punishes someone only if she aims to harm him. In a series of papers, Bill Wringe argues that this claim is false, I criticize his arguments, and he replies. Here, I argue that his reply fails.

Keywords Punishment · Harm · Intention

This paper continues a debate about the following claim.

> **The Aim to Harm Requirement [AHR]**: An agent punishes someone only if she aims to harm him.

In a series of papers, Bill Wringe argues that AHR is false (Wringe 2013), I criticize his arguments (Hanna 2017), and he replies (Wringe 2019). In this paper, I argue that his reply fails. This debate matters because the nature of punishment matters. Whether AHR is true is an inherently interesting and potentially morally significant question. Many philosophers of punishment accept AHR or claims like it and think that the truth of such claims makes punishment especially hard to justify.[1] If such claims are false, punishment may be easier to justify than these philosophers think.

Wringe and I disagree about a lot, so I'll have to limit my discussion to the most important points of disagreement between us. I'll start by sketching the standard argument for AHR and explaining why Wringe rejects the argument.

[1]Boonin says that such claims are "almost universally accepted" by philosophers of punishment (2008: 13–14, n14; cf. Hanna 2013: 593, n5). Those who think that such claims make punishment especially hard to justify include Berman (2008: 267), Boonin (2008: 12-17), Tadros (2011: 1, 265), and Zimmerman (2011: 7-10, 164).

✉ Nathan Hanna
nhanna@drexel.edu

[1] Drexel University, Department of English and Philosophy, MacAlister Hall, Room 5044, 3250-60 Chestnut Street, Philadelphia, PA 19104, USA

1 Preliminaries

The standard argument for AHR says that AHR is sometimes necessary to distinguish punishments from non-punishments. Here's a sketch of the argument. We can treat two people in comparably harmful ways for similar reasons and only punish one of them. For example, we can 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. AHR is necessary to distinguish punishments from non-punishments in such cases: typically, we're punishing only the culpable killer because we aim to harm him and not the non-culpable killer. So AHR is true.[2]

Wringe objects that collective agents like states can punish without aiming to harm and that other claims can make the relevant distinctions.[3] These other claims form part of his account of legal punishment, which he puts as follows.

> I am interested in giving an account of … legal punishment in political communities with modern legal systems. The most prominent examples of such communities … are states. … I take states and political communities more generally to be best understood as collective agents. I hold that the following conditions are necessary and sufficient for … punishment: 1) It should involve harsh treatment 2) Inflicted by an appropriate authority 3) On an offender 4) In response to some specifiable wrongdoing 5) With an appropriate expressive purpose. (Wringe 2019: 4–5)

"Harsh treatment" is Wringe's term of art for "treatment which would normally be found [harmful] by a typical individual of the kind on whom it is being imposed" (ibid.).[4] Because such treatment only has to be normally harmful, Wringe maintains that punishment doesn't have to be harmful or aimed at being so.

In my previous response to Wringe, I argued that his account's conditions don't suffice for punishment. I did this by giving counterexamples to the account – cases where it counts non-punishments as punishments (Hanna 2017: 970–71). Wringe replies by arguing that these cases aren't counterexamples and that there are apparent counterexamples to AHR. In the next two sections, I'll argue that Wringe is wrong on both points. Along the way, I'll also argue that his account can't obviously make the distinctions that the standard argument for AHR appeals to.

2 Counterexamples I

In my previous response to Wringe, I argued that the following case is a counterexample to his account (Hanna 2017: 970–71).

[2] For versions of this argument see Boonin (2008: 12-15), Hanna (2009: 330-31), Honderich (1969: 1), Lee (2019-363-64), Ten (1987: 15), Wasserstrom (1982: 476), and Zimmerman (2011: 9-10). I say "typically" because it's possible for an agent to punish someone that the agent knows isn't culpable and because psychiatric commitment can be used to punish (see section 3 below and cf. Hanna 2017: 974).

[3] There are many other objections to AHR and to the standard argument for it that I won't discuss. The most common ones misinterpret what AHR means by "aims" or "harm." For responses to these and other well-trodden objections see Boonin (2008: 6-17), Hanna (2014: 593-94), Lee (2019: 371, 377-79), and Zimmerman (2011: 3-10, 19-21).

[4] Wringe uses several synonyms for "harmful." I'll stick with "harmful" to avoid confusion.

> **Judgment**: Judge must sentence Thief, but she 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 to the two of them – won't harm him. He happens to love doing community service and was planning to do some. Judge sentences him to do the very kind that he was planning to do. He does even more than required, greatly enjoys it, and doesn't find it at all unpleasant.

I said that Thief isn't punished and that Wringe's account entails otherwise. All five of the account's conditions seem to be satisfied here (or will be if we add the right stipulations to the case). Notably, the sentence is harsh in Wringe's sense because people are almost always harmed by being sentenced to do community service.[5]

Wringe doesn't deny that his conditions are satisfied here. Instead, he challenges the intuition that Thief isn't punished. He says that the sentence arguably harms Thief because it restricts Thief's liberty and because "it's not implausible to think of liberty as an element of well-being" (Wringe 2019: 7). He takes this to cast doubt on the intuition that Thief isn't punished. (I'll discuss some other ways that he might reply to the case later).

Wringe is wrong about Judgment. Moreover, there are variants of the case that aren't vulnerable to his reply. I'll take these points in turn. He's wrong about Judgment because the following claim is true.

H Thief's liberty is harmlessly restricted.

You might object that H assumes a controversial theory of well-being.[6] But it doesn't. H is consistent with many different theories. And it's independently plausible. I'll explain.

The leading theories of well-being include desire satisfaction theories, hedonist theories, objective list theories, and perfectionist theories.[7] H is consistent with each type of theory. It's consistent with desire satisfaction theories because Thief's desires aren't frustrated. And it's consistent with hedonist theories because he's not caused pain or deprived of pleasure. As for the other theories, some of them take liberty to be an element of well-being and some don't. H is consistent with the latter theories because Thief isn't deprived of anything else that they take to be elements of well-being, e.g., achievement, knowledge, or virtue. H is also consistent with at least some of the former theories, e.g., those that say that a certain *amount* of liberty is an element of well-being.[8] This is because Thief still has lots of liberty. All of this shows that H is consistent with many different theories of well-being.

I also said that H is independently plausible. Here's an argument for it. Well-being seems prudentially important. It seems irrational for an agent to be indifferent to the fact that he's been made worse off in some way.[9] Yet it seems rational for Thief to be indifferent to the fact that his liberty has been restricted and to think that he got off scot-free. This is because of the way that his liberty has been restricted: he's been sentenced to do something that he was going to do anyway, loves doing it, and doesn't find it at all unpleasant. So H is true.

[5] Wringe seems to agree (Wringe 2019: 8, n19). As for the other conditions, 2–4 are clearly satisfied. And Wringe seems to think that 5 can be satisfied by background features of the legal system (ibid.: 6–7). We can stipulate that these features are present or opt for some stipulations that I suggested in my previous reply (Hanna 2017: 971, n7).

[6] Thanks to a referee.

[7] For overviews of these theories see Bradley (2015) and Crisp (2017).

[8] Kazez endorses such a theory (Kazez 2007: 65–68, 79).

[9] Cf. Bradley (2012: 395, 2015: 7-8), Crisp (2017: Sect. 1).

This argument appeals to some plausible intuitions about well-being – intuitions that Wringe's reply has to reject. That's a mark against his reply. Maybe you reject these intuitions, though. No matter. We can argue against Wringe's account without appealing to them. This brings me to the second problem with his reply to Judgment: there are versions of Judgment where Thief's liberty isn't restricted. Here's one.

> **Judgment***: As in Judgment, except as follows. Before being sentenced, Thief mailed $1000 in compensation to his victim. Only he and Judge know this. The money won't arrive until after he's sentenced and he can't now stop his victim from getting it. Judge sentences him to give $1000 in compensation to his victim.

This sentence doesn't restrict Thief's liberty because he already sent the money and can't stop his victim from getting it. The case is otherwise like Judgement. It seems that Thief isn't punished, but Wringe's account entails otherwise. All five of the account's conditions are satisfied here (or will be with the right stipulations). Notably: a $1000 compensation order is harsh in Wringe's sense. So the case seems like a counterexample to his account.

Wringe might reply by arguing that the sentence here has other harmful features. But it's hard to see what those could be. And anyway, we could just come up with other versions of Judgment that lack the allegedly harmful features. So even if Wringe's reply to Judgment had worked – and I've argued that it doesn't – it can't obviously generalize to every version of the case. We can come up with different versions of the case to suit different intuitions about well-being and harm. There's no reason to think that Wringe can evade every such counterexample.[10]

Wringe seems to have anticipated such a reply, though, because he goes on to argue that consulting our intuitions about hypothetical cases isn't a good way to assess the truth of his account or of AHR (Wringe 2019: 8–9). If he's right, the fact that his account conflicts with such intuitions isn't a problem for him. I'll discuss this argument later. For now, I'll discuss some cases that he takes to be apparent counterexamples to AHR. He thinks that they show that we should reject AHR even if he's wrong about the significance of our intuitions.

3 Counterexamples II

Here's a simplified version of Wringe's first case (cf. Wringe 2019: 8).

> **Incompetent Judge (simplified)**: As in Judgment, except for the following. Judge is mistaken about Thief's preferences and plans. He hates doing community service and grudgingly does only what he has to do.

[10] Cf. Hanna (2017: 971). Alternatively, Wringe might deny that the sentences in these cases are harsh because they can be described in unharsh ways. For example, he might say that the sentence in Judgment* isn't harsh because it can be described as "sentencing someone to do what they've already done," which isn't harsh (thanks to a referee). This reply doesn't obviously work. Recall, Wringe thinks that we can punish someone harmlessly, so long as we treat her harshly. In such cases, we can describe the treatment in an unharsh way: whatever feature of the case makes the treatment harmless can be used to describe it that way. To make the reply work, Wringe would need principled reasons to use descriptions that get his desired results in these cases and that avoid the problematic results in mine. He gives none. And, as I argued in my previous reply to Wringe, it's not obvious that there are any (Hanna 2017: 971–72).

Wringe says that Thief is punished here and that AHR entails otherwise. But Wringe is wrong about what AHR entails. AHR says that an agent punishes someone only if she aims to harm him. This doesn't entail that Thief isn't punished – it just entails that Judge doesn't punish him. Wringe actually accepts the latter claim.[11] Since my purpose here is just to respond to him, I won't defend it. That said, those who think that Judge does punish Thief in this case can accept a variant of AHR: an agent punishes someone only if she aims to harm him or acts on behalf of someone who does. The case isn't a counterexample to this variant.[12]

Wringe's second case might threaten both AHR and its variant, though. Here's a simplified version of the case (cf. Wringe 2019: 9).

> **Stoic State (simplified):** In Politeia, laws are made and enforced by stoics who mistakenly believe that people are harmed only if they become less virtuous, that people are benefitted only if they become more virtuous, and that imprisoning offenders makes them more virtuous, not less. The stoics want to benefit and avoid harming offenders. So they create and administer laws and institutions to imprison offenders.

Wringe says that the offenders here are punished and that AHR entails otherwise. The case lacks crucial details, though. I'll argue that, depending on how we flesh out the details, the claim that the offenders aren't punished is either plausible or not entailed by AHR. Showing this will require an extended discussion, so bear with me.[13]

Consider what AHR entails. Wringe thinks that it entails that the offenders aren't punished, presumably because the stoics aim to avoid inflicting harm *as they understand it*. This reasoning takes AHR to be saying that an agent aims to harm someone only if she believes that what she's aiming at is harmful. But that's not a good way to understand AHR. Whatever an agent believes about harm, it's plausible to say that she aims to harm someone if she satisfies either of the following conditions.

- She aims to subject him to something and that thing is intrinsically bad for him – candidates include pain, suffering, and frustration.
- She aims to deprive him of something and that thing is intrinsically good for him – candidates include pleasure, happiness, and satisfaction.

In other words: whatever an agent believes about harm, it's plausible to say that she aims to harm someone if she aims to do things like make him suffer or deprive him of happiness.[14] Alice Ristroph makes a similar point about related concepts.

[11] He thinks that, in modern legal systems, punishment is inflicted by collective agents like states and political communities, not by individual officials like judges (Wringe 2019: 4, 7).

[12] Zimmerman proposes this variant (2011: 19–20). In my previous reply to Wringe, I said that Zimmerman's variant can deal with such cases (Hanna 2017: 973, n12). Wringe misunderstands this part of my discussion (Wringe 2019: 7, n15). For the record, I'm inclined to accept AHR as is. Even if Thief is punished, I'd say that Judge doesn't punish him – she just unwittingly helps others to do so. I won't defend this claim here.

[13] Ristroph describes a somewhat similar case that can also be used to argue against AHR (Ristroph 2008: 1399–1400). I don't have the space to do justice to her case here, but I'd say similar things about it.

[14] The above conditions may need to be qualified in certain ways, but I won't try to refine them here (thanks to a referee for pressing me on this point). Also, note that these are just sufficient conditions. There may be other sufficient conditions. I take no stand on what the full set of necessary and sufficient conditions is.

> A criminal statute may prohibit 'intentional maiming.' A defendant swings a stick intending to put out his victim's left eye, but the concept of *maiming* does not cross [his] mind. Even though [he] does not formulate his own purpose in terms of maiming per se, a fact-finder might reasonably conclude that intent to put out an eye is properly classified as intent to maim. (2008: 1397)

So understood, AHR doesn't entail that the offenders in Stoic State aren't punished. This is because the case doesn't say if the stoics are aiming to do things like make offenders suffer or deprive them of happiness. The case doesn't say if the stoics are aiming to do such things because it doesn't say what features of imprisonment they take to promote virtue. For all that the case says, the stoics may be aiming to make offenders suffer in the belief that this will make offenders more virtuous.

I've argued that AHR doesn't entail that the offenders in Stoic State aren't punished. However, there may be ways to frame the case so that AHR does entail this. To see if this is a problem, suppose that AHR entails this if we stipulate that the stoics don't aim to do things like make offenders suffer or deprive them of happiness. Call this case Stoic State*. I think that AHR's implications here are plausible. To see why, recall the standard argument for AHR.

> We can treat two people in comparably harmful ways for similar reasons and only punish one of them, e.g., we can 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. AHR is necessary to distinguish punishments from non-punishments in such cases. So it's true.

Here's a similar argument that Stoic State* isn't a case of punishment: If it is, there must be a relevant difference between it and the non-culpable killer case that can explain why; there's no such difference; so it's not a case of punishment. The first premise of this argument is undeniable. So you'll have to deny the second premise if you find AHR's implications here counterintuitive. I'll defend this premise by showing that several objections to it fail. Similar reasoning can be used to deal with other alleged counterexamples to AHR. My reasoning will also lend indirect support to the standard argument for AHR – for each objection that I'll discuss, there's an analogous objection to the standard argument that fails for the same reasons.

Let's start with this objection.

- Imprisonment is a form of punishment and psychiatric commitment isn't. Stoic State* involves the former and the non-culpable killer case involves the latter.

This objection fails. For one thing, people can be imprisoned without being punished. Pre-trial detention illustrates this. For another thing, psychiatric commitment can be used to punish. A case can illustrate.[15] Suppose that a small town's authorities – including the prosecutor's office, police force, and judiciary – are convinced of a defendant's guilt. They want to harm him severely for his crime, but the law allows only a mild sentence for it. So they conspire to misuse another part of the law that lets them indefinitely commit legally incompetent defendants in psychiatric institutions. Some of the officials dishonestly raise doubts about the defendant's competence, others solicit or supply false testimony supporting a finding of incompetence, and the judiciary abets it all, ultimately finding him incompetent and

[15] Thanks to a referee for pressing me to add details to this case.

committing him indefinitely. It seems clear that the defendant here is punished by being committed in a psychiatric institution. So the above objection fails.

These remarks might suggest some better objections to my argument that Stoic State* isn't a case of punishment, though. Here are two that draw on Wringe's account of legal punishment and that mirror his objections to the standard argument for AHR (cf. Wringe 2013: 869–70).

- Legal punishment is a response to wrongdoing. Stoic State* involves a response to wrongdoing and the non-culpable killer case doesn't.
- Legal punishment has an appropriate expressive purpose. Stoic State* involves such a purpose and the non-culpable killer case doesn't.

These objections don't identify a relevant difference between the cases. The first doesn't because the non-culpable killer does wrong – he just has an excuse.[16] The second doesn't because it doesn't say what an appropriate expressive purpose is. Committing the non-culpable killer in a psychiatric institution can have many expressive purposes, e.g., it can express disapproval of killing, sympathy for victims, and reassurance to the public. For all that the second objection says, any of these can be an appropriate expressive purpose.[17] So the above objections also fail.

Maybe we can improve on the previous objections as follows, though.

- Legal punishment is a response to culpability. Stoic State* involves a response to culpability and the non-culpable killer case doesn't.
- Legal punishment expresses moral condemnation. Stoic State* involves an expression of moral condemnation and the non-culpable killer case doesn't.[18]

The first objection fails because legal punishment isn't necessarily a response to culpability. Punishments for strict liability offenses show this.[19] As for the second objection, there are two ways to interpret its claim about legal punishment. One: legal punishment morally condemns the punishee. Two: it morally condemns the punishee's actions. Claim one is false. Punishments for strict liability offenses show this. And claim two doesn't identify a relevant difference between the cases. Committing the non-culpable killer in a psychiatric institution can morally condemn his actions as harmful, unjust, or cruel say. So these objections also fail.

Consider one last objection, then. It draws on another part of Wringe's account, though I don't know if he'd endorse it.

[16] Cf. Husak (2008: 76). In my previous reply to Wringe, I argued that legal punishment isn't necessarily a response to wrongdoing (Hanna 2017: 974–75).

[17] Compare these remarks of Wringe's: "Can we say anything general about the kinds of expressive purpose that are essential to punishment? Not much, I suspect. Clearly not any expressive purpose will do… There must at least be reasons for taking the harsh treatment to express something about the wrongdoing (or [about] the character or motivations of the [wrongdoer]" (2019: 5; cf. Feinberg 1965: 402–3).

[18] Wringe would likely endorse the first objection because he endorses a similar objection to the standard argument for AHR (cf. Wringe 2013: 870). He'd reject the second objection because he denies the moral condemnation claim (Wringe 2019: 5–6) For arguments against the claim see Hanna (2017: 974-75) and Lee (2019).

[19] To clarify, I'm not saying that strict liability punishments are never responses to culpability. In many such cases, even though we don't require evidence of culpability for conviction, we might still know or at least believe that the punishee is culpable. My point is just that it's possible to impose strict liability punishments on people who we know aren't culpable. We might do this to enhance deterrence, say. Thanks to a referee for pressing me here.

- Legal punishment is inflicted on offenders. Stoic State* involves an offender and the non-culpable killer case doesn't.

This objection doesn't identify a relevant difference between the cases either. The non-culpable killer can still be an offender – i.e., someone who violated the criminal law – because he may very well have satisfied the actus reus and mens rea requirements for an offense.[20] Moreover, the objection says something false about legal punishment: it's not necessarily inflicted on offenders. To see this, consider a case. A low level government employee is also a prominent political dissident. Her political activity is entirely legal and hasn't interfered with her job. But it has angered lawmakers. They want to punish her for it, but don't want to use the criminal law to do so. Instead, they pass a law barring her from government employment. Their aim is to harm her for what they see as subversive political activity. This seems like a form of legal punishment because the legislators are using their legislative powers in order to inflict what seems to be a punishment.[21] So this objection fails too.

I've considered several objections to my argument that Stoic State* isn't a case of punishment. They all fail, either because they say false things about punishment or because they don't identify a relevant difference between Stoic State* and the non-culpable killer case. I suspect that other objections would fail for the same reasons, though I obviously haven't shown this since I haven't considered every possible objection. That said, the failure of the above objections casts considerable doubt on the claim that Stoic State* is a case of punishment. Their failure suggests that intuitions for this claim lack a good supporting explanation and are based on a superficial assessment of the case. I therefore conclude that Stoic State* isn't obviously a counterexample to AHR. And because Incompetent Judge and Stoic State aren't counterexamples either, I conclude that Wringe hasn't obviously given a counterexample to AHR.

Wringe actually agrees that these aren't counterexamples, though. This is because, as you'll recall, he thinks that consulting our intuitions about hypothetical cases isn't a good way to assess the truth of his account or of AHR. His cases are just meant to show that we should reject AHR even if he's wrong about this. He gives another argument against AHR that doesn't rely on these intuitions. I'll conclude this section by discussing this argument. Here's how he puts it.

[20] Objection: the fact that he's non-culpable due to mental illness entails the absence of mens rea (thanks to a referee). Reply: this is false. The law often defines mens rea in terms that even non-culpable mentally ill defendants can satisfy (cf. §2.02 and §4.01 of the Model Penal Code). As Morse and Hoffman put it: "Crimes are defined by their 'elements,' which always include a prohibited act and in most cases a mental state, a mens rea, such as intent… Even if the state can prove all the elements beyond a reasonable doubt, the defendant may avoid criminal liability by establishing an affirmative defense of justification or excuse" (2007: 1074). They go on to give examples where insanity defenses were successful despite the presence of mens rea, so understood: "The examples of Daniel M'Naghten and Andrea Yates … demonstrate that even the most delusional or hallucinating person can form the requisite mental state. M'Naghten … surely intended to kill a person. Likewise, Andrea Yates … surely intended to kill [her] five children" (ibid.: 1089–90).

[21] US courts recognize the possibility of such punishment and interpret the Constitution's ban on bills of attainder as prohibiting it. For an overview of relevant case law see Dick (2011). My case is modeled on the one in *United States v. Lovett*, 328 U.S. 303 (1946). According to the Lovett court, "Legislative acts, no matter what their form, that apply … in such a way as to inflict punishment on [people] without a judicial trial, are bills of attainder prohibited by the Constitution." See also *United States* vs. *Brown*, 381 U.S. 437 (1965): "The Bill of Attainder Clause … was intended to implement the separation of powers among the three branches of the Government by guarding against the legislative exercise of judicial power [and] is to be liberally construed in the light of its purpose to prevent legislative punishment." Thanks to a referee for pressing me here.

> I don't take [cases like Incompetent Judge and Stoic State] to show [that AHR is false]… I take the falsity of AHR to be demonstrated by the fact that when we impose harsh treatment on an offender but fail to harm them we don't take ourselves to have failed to punish them (and our punitive practices bear this out). (Wringe 2019: 9, n22)

In this passage Wringe seems to be arguing that AHR is false because there are real-world cases of harmless punishment. The problem is that he doesn't give any good examples to support the latter claim.[22] He says that our punitive practices show that it's true, but he doesn't say which practices he has in mind.[23] This wouldn't be a problem if there were obvious and uncontroversial cases of harmless punishment. But there aren't.[24] So his argument fails.

In the next and final section, I'll consider Wringe's argument for the claim that consulting our intuitions about hypothetical cases isn't a good way to assess the truth of his account or of AHR. I'll also consider another argument of his: that my views about punishment's justifiability give me reasons to reject AHR.

4 Methods and Morals

Wringe thinks that consulting our intuitions about hypothetical cases isn't a good way to assess the truth of his account or of AHR. This is partly why he denies that cases like Judgment are counterexamples to his account. His reasons for doubting these intuitions are worth quoting at length.

> Hanna's counter-examples appeal to what we might call a folk conception of punishment. If they had succeeded, they would – at best – have shown that conditions 1-5 did not perfectly correspond to our folk intuitions about punishment. However, it's not obvious to me that philosophical theorizing about punishment should be concerned with giving an analysis of a folk conception of punishment. …[I'm] interested in giving an account of the phenomenon of punishment in political communities with modern legal systems. The question of how such a conception might relate to our folk conception of punishment is not straightforward. One possibility is, of course, that the relationship is one of identity. Another, is that 'punishment' is a theoretical kind term of intellectual enterprises such as criminology and penology. On this view, we might expect a theoretical conception of punishment to be revisionary of some of our intuitions about punishment, just as a theoretical account of water might lead us to revise some of our folk judgments about what counts as water. If so, then even if Hanna's counter-examples established what he thinks they establish, it's not clear how worried an expressivist should be. (2019: 9)

[22] He cites a section of his earlier paper where he discusses some hypothetical cases that he seems to think are cases of harmless punishment (Wringe 2013: 867). But since they're hypothetical, they don't support his premise. And they're not obviously cases of harmless punishment anyway, as I argued previously (Hanna 2017: 971, n8).

[23] His talk of failure is suggestive, though. The sections of his 2013 paper that he cites contain what seems to be an argument against AHR that appeals to the claim that harmless attempts at punishment are failures if AHR is true. Lee shows that this argument fails (Lee 2019: 371–72, 379 n57).

[24] Boonin argues that many of those who think that there are cases of harmless punishment misunderstand the concept of harm (Boonin 2008: 6–17; cf. Hanna 2014: 593). Wringe does this at points (cf. Hanna 2017: 971, n8).

In this passage, Wringe argues that my counterexamples exploit an intuitive conception of punishment that might misrepresent the nature of modern legal punishment. In defense of this claim, he says that some of the social sciences might be using a different and more accurate conception. There are two problems with this argument.

First, Wringe doesn't show that the relevant sciences use a different conception of punishment. Second, even if they do, he gives no reason to think that such a conception might be more accurate. Suppose for argument's sake that some of the social sciences do use a different conception. That's no reason to doubt our intuitive conception. Even if part of our intuitive conception is missing from the social scientific conception, that part of our intuitive conception might still be true. It might be missing from the social scientific conception just because it's irrelevant to certain kinds of social scientific work, not because it's false. AHR might be like that. It might be true but irrelevant to certain kinds of social scientific work like attempts to assess the deterrent effects of punishments such as imprisonment or the death penalty. AHR might be more relevant to philosophical rather than scientific work, e.g., attempts to answer foundational questions about the nature and ethics of punishment.

This reply might seem to overlook something important, though. At one point Wringe seems to suggest that a social scientific conception of punishment might be more accurate, not just because it's used in the sciences, but because it might be informed by scientific work. He seems to suggest this when he says that a social scientific conception might lead us to revise our intuitive conception just like work in the hard sciences led us to revise our folk conception of water. But this is a bad analogy. The claim that cases like Judgment aren't cases of punishment isn't the sort of claim that science can disconfirm. Scientific methods can't be used to test its truth any more than they can be used to test the truth of analogous claims about free will or harm, e.g., that compatibilism is true or that a comparative account of harm is true.[25]

To sum up, Wringe doesn't given any good reasons to doubt our intuitive conception of punishment. He doesn't give any good reasons to think that intuitions about hypothetical cases aren't a good way to assess the truth of his account or of AHR. And he doesn't give any good reasons to reject my counterexamples to his account.

By way of conclusion, I want to consider one more argument of Wringe's. It concerns AHR's moral significance. Some philosophers think that AHR or claims like it support abolitionism, the view that legal punishment is unjustified. I'm one of them.[26] Wringe argues that abolitionists should reject AHR. I'll quote him at length again.

> Considerations about abolitionism provide a further reason for countenancing a revisionary approach to our folk conception of punishment. Suppose that Hanna is right and that AHR is a component of our folk conception of punishment. There could, nonetheless be institutions very like our own that … did not involve an intention to cause [harm]. Call such treatment 'wunishment' and such institutions 'wunitive'. Wunitive institutions might be – as Hanna suggests – better in some respects than our existing institutions: they might avoid deliberate cruelty, and bring about less suffering. But it seems at least as plausible that many of the morally unattractive features of our existing

[25] A qualification: I take no stand on whether the sciences can test such claims by probing our intuitions. Wringe doesn't cite any empirical research purporting to show that we have intuitions about punishment that are inconsistent with AHR.

[26] See, e.g., Boonin (2008: 16-17, 213, passim), Hanna (2009: 331, 2014: 592-95), Zimmerman (2011: 164-5). Cf. Sayre-McCord (2001: 503-10).

> punitive institutions would be features of wunitive institutions too. In particular, it is hard to see why they could not be as stigmatizing, degrading and dehumanizing as the worst existing American prisons are. On Hanna's view of punishment, a proposal to replace institutions that are punitive with ones that are … merely wunitive ought to satisfy the abolitionist's demands. But this would be abolitionism on the cheap. It ought not to satisfy those whose adoption of abolitionism is motivated by an awareness of the awfulness of these features of American prisons. (2019: 9–10)

This misunderstands what abolitionists like me think. I'll explain.

Abolitionists argue that the aim to harm makes legal punishment unjustified. This doesn't mean that they take the aim to be the only problem – or even the worst problem – with our punishment practices. Many philosophers of punishment think that these practices are bad in lots of ways, e.g., that we criminalize too many things and inflict too much harm.[27] Abolitionists agree. They focus on the aim to harm because they think that there's something special about the problem that it poses, not because they don't appreciate these other problems. Most philosophers of punishment seem to think that we can fix these other problems by reforming how we punish, e.g., by decriminalizing many things and reducing sentence severity. They think that legal punishment is justified in principle. Abolitionists agree that such reforms are needed – they just think that such reforms don't go far enough. They think that legal punishment is unjustified in principle because they think that an essential feature of it makes it unjustified: the aim to harm. If they're right, the only way to fix this problem is to abolish legal punishment.[28] That's why they focus on the aim to harm.

You might object that this reply misses a larger point, though. Wringe seems to think that the aim to harm isn't obviously a problem with legal punishment even if AHR is true – or at least not as serious of a problem as abolitionists think. This seems to be the point of his claim that "wunishment" could be just as harmful as the worst contemporary forms of legal punishment. In light of this, you might say that abolitionists have to explain why they think that the aim to harm is so morally important. The problem with this complaint is that they've given a number of explanations for this view.[29] I won't discuss these explanations here because Wringe doesn't engage with them.[30] Moreover, even if these explanations are false and abolitionists are mistaken about the moral importance of the aim to harm, that wouldn't have any obvious bearing on the truth of AHR. AHR might be true even if it's not as morally important as abolitionists think. Wringe's claims about the moral importance of the aim to harm are just irrelevant to the debate about whether AHR is true.

I conclude that Wringe hasn't given any good reasons to think that abolitionists like me should reject AHR. I further conclude that his account of legal punishment is false and that he gives no good reasons to reject AHR.[31]

[27] See, e.g., Husak (2008), Wellman (2017: 172-191), Tadros (2011: 2).
[28] They typically argue that there are viable alternatives to legal punishment that don't aim to harm. See, e.g., Boonin (2008: 213-75) and Hanna (2009: 332-47; 2014: 600-602). Cf. Sayre-McCord (2001: 503-10).
[29] Boonin (2008: 15-17, 234), Hanna (2009: 331; 2014: 595-97), Sayre-McCord (2001: 507), Zimmerman (2011: 165-75).
[30] I discuss them elsewhere, identify some problems with them, and offer a better explanation (Hanna ms).
[31] Thanks to Stephen Galoob and to two referees for comments.

References

Berman M (2008) Punishment and justification. Ethics 118:258–290
Boonin D (2008) The problem of punishment. Cambridge University Press, New York
Bradley B (2015) Well-Being. Polity Press, Cambridge
Bradley B (2012) Doing away with harm. Philos Phenomenol Res 85:390–412
Crisp R (2017) Well-being. The Stanford Encyclopedia of Philosophy, Edward N. Zalta (ed.), https://plato.stanford.edu/archives/fall2017/entries/well-being/
Dick A (2011) The substance of punishment under the bill of attainder clause. Stanford Law Review 63:1177–1211
Feinberg J (1965) The expressive function of punishment. Monist 49:397–423
Hanna, N. (ms) Hearts of Darkness: Why Punitive Intent Matters
Hanna N (2017) The nature of punishment: reply to Wringe. Ethical Theory Moral Pract 20:969–976
Hanna N (2014) Facing the consequences. Crim Law Philos 8:589–604
Hanna N (2009) Liberalism and the general justifiability of punishment. Philos Stud 145:325–349
Honderich T (1969) Punishment: the supposed justifications. Harcourt, Brace & World, New York
Husak D (2008) Overcriminalization. Oxford University Press, New York
Kazez J (2007) The weight of things. Blackwell Publishing, Oxford
Lee AYK (2019) Arguing against the expressive function of punishment: is the standard account that insufficient? Law Philos 38:359–385
Morse SJ, Hoffman MB (2007) The uneasy entente between insanity and Mens Rea: beyond *Clark v. Arizona*. The Journal of Criminal Law and Criminology 97:1071–1150
Ristroph A (2008) State intentions and the law of punishment. The Journal of Criminal Law and Criminology 98:1353–1406
Sayre-McCord G (2001) Criminal justice and legal reparations as an alternative to punishment. Philosophical Issues 11:502–529
Tadros V (2011) The ends of harm: the moral foundations of criminal law. Oxford University Press, Oxford
Ten CL (1987) Crime, guilt, and punishment. Clarendon Press, Oxford
Wasserstrom R (1982) Capital punishment as punishment: some theoretical issues and objections. Midwest Studies in Philosophy 7:473–502
Wellman CH (2017) Rights forfeiture and punishment. Oxford University Press, New York
Wringe B (2019) Punishment, judges, and jesters: a response to Nathan Hanna. Ethical Theory Moral Pract 22:3–12
Wringe B (2013) Must punishment be intended to cause suffering? Ethical Theory Moral Pract 16:863–877
Zimmerman M (2011) The immorality of punishment. Broadview Press, Peterborough