

Prevention and Imminence, Pre-Punishment and Actuality

GIDEON YAFFE*

TABLE OF CONTENTS

I.	INTRODUCTION	1205
II.	PUNISHMENT AND THE ACTUALITY OF THE CRIME.....	1208
III.	PREVENTION AND THE IMMINENCE OF THE CRIME.....	1217
IV.	CONCLUSION	1226

I. INTRODUCTION

In a variety of circumstances, it is justified to harm persons, or deprive them of liberty, in order to prevent them from doing something objectionable. We see this in interactions between individuals—think of self-defense or defense of others—and we see it in large-scale interactions among groups—think of preemptive measures taken by countries against conspiring terrorists, plotting dictators, or ambitious nations. We can argue, of course, about the details. Under exactly what conditions is it justified to inflict harm or deprive someone of liberty for reasons of prevention? But in having such arguments we agree on the fundamental idea: there are conditions under which prevention is justified, even if it is unclear precisely what those conditions are.

* Thanks to Adam Kolber and Jacob Ross for helpful written comments on an earlier draft. And thanks to Larry Alexander, Richard Arneson, Andrew Ashworth, David Brink, Kim Ferzan, Doug Husak, Ken Levy, Stephen Morse, Sam Rickless, Chris Slobogin, Steven Smith, Alec Walen, and Lucia Zedner for helpful discussion.

By contrast, it appears that it is never justified to punish in advance. “Pre-punishment,” as such a practice has been called,¹ seems no better than punishing for something the person did not do. Because punishment of the innocent is so abhorrent, the first inquiry in any criminal trial is an inquiry about actual innocence: assuming there was a crime, is the defendant the very person who committed it? It does not do to show that the defendant will commit the crime in the future or commit it if not punished. Pre-punishment is no better than punishment of the innocent.

Pre-punishment admits of two forms: defendants can be punished for something that they do, in fact, go on to do, or the punishment can cause the defendants not to do it and thus avert the crimes for which the defendants are punished.² One might think that although both kinds of pre-punishment are morally objectionable, the second is less bad: after all, at least the punishment in such a case serves to prevent something bad from happening. But in fact, from a retributivist point of view, to avert a crime through pre-punishment is actually worse than to punish in advance for a crime that actually occurs. When the crime is committed, after the fact, then the punishment serves to give the defendant something he or she deserves, albeit not something deserved at the time of the pre-punishment. The later commission of the crime might be thought to make an honest punishment, as it were, out of the pre-punishment.³ By contrast, a pre-punishment that causes the crime not to be committed is not only issued to someone undeserving at the time of the punishment; worse, that person never even becomes deserving of the punishment. In this respect, pre-punishment that averts crime differs not at all from punishment of the innocent: in both cases there is a punishment of a person for a crime never committed.

Prevention bears a greater similarity to pre-punishments that avert crime—those that are *worse* from a retributivist point of view—than it does to the other sort. After all, preventions, when effective, succeed in preventing those bad acts that they were issued to prevent. We therefore have a pair of puzzles. First, why are preventions justified under certain conditions when pre-punishments are not? Second, why are preventions that fail to avert crime harder to justify than preventions that succeed,

1. The term may have been coined in Christopher New, *Time and Punishment*, 52 ANALYSIS 35, 35–40 (1992).

2. There are also cases in which the defendant is punished in advance and does not commit the crime later, but the pre-punishment is not the cause of this fact. Set this group of cases aside.

3. We can imagine a retributivist who denies this by insisting that a punishment is not deserved if it does not follow the crime. In fact, much of what follows here is intended to explain what would justify such a retributivist position.

while by contrast, pre-punishments that do not avert crime are easier to justify than those that do?

This Article aims to solve this pair of puzzles. However, the Article has another aim, also, which some will see as more pressing and important. The other aim of the Article is to show that in a significant class of cases, the justificatory argument in favor of harming another in order to prevent that person from doing something objectionable is significantly weaker if that person's objectionable conduct is not imminent. This is so, it will be suggested, even if you are entirely certain that the person will do the objectionable thing if not prevented. Imminence is important to the justification of prevention, that is, independently of necessity. The argument for this controversial claim emerges from the discussion of the similarities between prevention and pre-punishment. Pre-punishment is wrongful, it is suggested, because the crime has not been actualized at the time of the punishment. A next best alternative to actualization is imminence. And so imminence aids in the justification of prevention for reasons closely related to the reasons why actualization is needed for the justification of punishment.

Part II concerns pre-punishment and identifies the sense in which punishment's justification depends on the actuality of the crime for which punishment is issued. The Part also justifies this actuality requirement on punishment through appeal to a necessary condition on desert of punishment—one found, among other places, in the writings of Samuel Pufendorf.⁴ Part III explains the sense in which imminence can serve not as a replacement for actuality but as a next-best alternative that allows for the justification of a certain class of preventions, namely, those in which part of what justifies the act of prevention is that the person thereby harmed would have deserved such harm had that person gone through with the objectionable act that the harm prevented. The Part also explains why imminence is of independent importance from necessity. The conclusion of the Article solves the two puzzles about the relationship between prevention and pre-punishment by drawing on the conceptual resources developed in the preceding Parts.

4. 2 SAMUEL PUFENDORF, *DE JURE NATURAE ET GENTIUM LIBRI OCTO* [ON THE LAW OF NATURE AND NATIONS] 91 (C.H. Oldfather & W.A. Oldfather trans., Clarendon Press 1934) (1688).

II. PUNISHMENT AND THE ACTUALITY OF THE CRIME

In a well-known paper, Christopher New defends the justice of pre-punishment in cases in which the punished persons go on to commit the crimes for which they are punished in advance.⁵ New offers a hypothetical in support of his claim. The hypothetical has three distinctive features: (1) it is known to all parties that a person is to commit a crime in the future, (2) if not punished in advance, the person never will be punished for it, and (3) the punished and the punisher explicitly agree to the terms of the punishment.⁶ New argues that in such cases there can be no principled objection—not from retributivism, not from deterrence theory—to punishment in advance.⁷ It is important to note that features (2) and (3) in New's example might distort intuition in ways that make it difficult to assess the relevance of the example to the conclusion that New seeks, namely, that pre-punishment per se is not objectionable. After all, it is possible that the general principle barring pre-punishment allows for exceptions, or can be overridden, by the inability to otherwise do justice, or by explicit or tacit agreement. If this were so, it would still at least be true that there is a pro tanto reason against pre-punishment.

However, putting aside concerns of this sort, New's paper raises a puzzle: what is it, exactly, about pre-punishment that is objectionable, even if there is only a pro tanto reason against it? In his discussion of the claim that retributivism is unable to explain this, part of New's argument relies on an analogy between punishment and payment for goods.⁸ It is perfectly legitimate to pay for goods either in advance or after receipt, depending on the contingent customs and tacit and explicit agreements governing such transactions. New's idea is that for the same reasons, it can be legitimate to "pay" for a crime by being punished, either in advance or after performance, depending on the nature of the contingent customs and tacit and explicit agreements governing punishment.⁹ If this is right, what follows is that there is nothing in the *nature* of punishment that makes it wrong to punish in advance. Or put another way, there is no convention- or agreement-independent moral reason not to punish in advance. We can imagine, for instance, different cultures, or different legal systems, settling the issue in different ways, and without differences in justice.

New's argument fails, but it fails constructively. It fails because there is a breakdown of the analogy between payment for goods and punishments

5. New, *supra* note 1, at 35–40.

6. *Id.* at 35–36.

7. *See id.* at 36–37.

8. *Id.* at 37.

9. *See id.* at 37–38.

for crimes. But there is an important lesson to be learned from the breakdown. To see this, first note that not even New would say that it could be just to pre-punish for a crime that the agent does not go on to commit. *That* brand of pre-punishment really is no different from punishment of the innocent. But why? After all, assuming that the punishment averts the crime—part of the reason that the punished person fails to commit it, for instance, is because of detention in prison at the time at which he or she would have committed it—then there is a nearby nonactual possible world in which the person commits the crime, namely, the world in which the person is not punished. Why should the existence of a nonactual possible world in which an agent commits a crime not suffice to justify punishment for that crime in the actual world? Why should we think that the crime and the punishment need both occur in the actual world for the punishment to be justified? There is no doubt that we do think this, but why do we think it?

Let us tag this undeniable principle with a name, conceding, for now, that it is not obvious why it is true:

The Actuality Principle: a punishment of person *D* for crime *C* is just only if *D* actually commits crime *C*.

The Actuality Principle is merely a restatement of the claim that it is never acceptable to punish individuals for crimes they do not commit. Notice—and this is the problem with New’s argument—there is no actuality constraint on payment for goods; it can be perfectly appropriate to pay for goods when the receipt of those goods is yet to be actual. By contrast, under the Actuality Principle, counterfactual commission of a crime will never do for actual punishment, even if the possible world in which such commission takes place is quite close to the actual world.

Now there are a variety of views that one might take about the actuality, or nonactuality, of both the future and the past. For instance, one might hold one of two symmetrical positions, positions that treat the past and the future as on an ontological par. So called presentists, for instance, think that the only thing that is actual is the present.¹⁰ On that view, neither the past nor the future is part of the actual world. Presentists must deny the Actuality Principle or else deny that punishment is ever just. After all, at the time of the punishment, the crime, even a past

10. For an extended defense of presentism, see generally CRAIG BOURNE, *A FUTURE FOR PRESENTISM* (2006).

crime, is not actual, and so under the Actuality Principle, the punishment for it cannot be just.¹¹ Presentists, then, must explain our belief that punishment of the innocent is wrong through appeal to some other fact about the innocent besides the nonactuality of the person's commission of the crimes for which punishment is sought. On another symmetrical view, so called four-dimensionalism,¹² the future, the past, and the present are all actual. On this view, differences in time are no more ontologically significant than differences in place; just as things that are in different places can both be in the actual world, so things at different times both inhabit the actual world. From a four-dimensionalist point of view, pre-punishment does not involve a violation of the Actuality Principle: future crimes are just as actual as past crimes, and so punishment in the present for a four-dimensionalist would not be a violation of the Actuality Principle. If there is something wrong with pre-punishment, the four-dimensionalist would have to explain it some other way.

Symmetrical views need to say something to accommodate a particular article of common sense, namely, that there is some kind of ontological asymmetry between the future and the past. The past is "fixed" or "given" or "part of the world" in a way that the future is not. Symmetrical views can accommodate this intuition by insisting that it is not an intuition about what is actual but some other kind of intuition. Perhaps it is an intuition about what depends on our present conduct; the past does not, though the future might. However, it is important to see that to adopt such an approach is to accept an error theory: we typically think that the future is not actual but the past is. Presentists and four-dimensionalists must deny that the intuition is true, and so they must paint ordinary thought as in error. On another view, call it the asymmetry view, the intuition is veridical: the past, like the present, is actual, but the future is not; it will become actual as time passes.¹³ If this is right, then pre-punishment violates the Actuality Principle, although punishment after the fact does not.

None of these three views is without defenders among metaphysicians. And given the Actuality Principle, it matters, for the reasons just given, which one is true. However, that question will not be settled here and, in fact, need not be. The crucial point for our purposes here concerns, instead, the rather complex relationship between moral intuition, metaphysical agnosticism, and commonsense intuition. Most of us have no idea if the

11. Perhaps under the presentist view, the only time at which punishment of a crime is just is the moment at which it is taking place.

12. Four-dimensionalism is defended in THEODORE SIDER, *FOUR-DIMENSIONALISM: AN ONTOLOGY OF PERSISTENCE AND TIME* (2001).

13. The asymmetry view, sometimes also called the Growing Block view, is defended in MICHAEL TOOLEY, *TIME, TENSE, AND CAUSATION* 116–17 (1997).

future is actual; we have the intuition that it is not, but we do not know enough about the arguments for and against symmetrical and asymmetrical views of actuality to know whether that intuition is true. Given our ignorance, to pre-punish is to risk the possibility of violation of the Actuality Principle; for all we know, after all, the asymmetry view is correct. Reasonable people might think it is true, and so there is reasonable doubt about the justice of pre-punishment simply because such punishment might be in violation of the Actuality Principle. For all we know, that is, it is just as wrong to pre-punish even when there will be later crime for which the punishment is fitting as it is to punish for a crime that only took place in a nonactual possible world.¹⁴

Now, one might worry that the same form of metaphysical agnosticism will rule out just punishment for past wrongdoing. After all, we might say, presentism might be true, and if it is, then punishment for past wrongdoing is in violation of the Actuality Principle. But here the fact that common sense tells us that the past is actual and the future is not holds sway. In other words, metaphysical agnosticism influences and ought to influence moral intuition when all of the competing metaphysical positions are consistent with common sense. But common sense silences doubts that have their source in metaphysics when there is a metaphysical position about which one is agnostic that is consistent with common sense. In short, we ought not to pre-punish because presentism or the asymmetry view might be true, and common sense does not tell us that they are not. But it is acceptable to punish after the fact because common sense supports the claim that the past is actual.

New claims that intuitions, to the effect that there is something wrong with pre-punishment, arise from epistemological sources.¹⁵ His idea is that we think it wrong to pre-punish because we worry that the punisher does not know—or as we might say, does not *really* know—that the

14. Here we have a compatibilist response to Saul Smilansky's assertion that compatibilists cannot explain what is wrong with pre-punishment. Saul Smilansky, *Determinism and Prepunishment: The Radical Nature of Compatibilism*, 67 ANALYSIS 347–48 (2007). Smilansky makes the mistake of thinking that it follows from the fact that a future event is causally determined—guaranteed to occur given the state of the present and the laws of nature—to be actual despite being future. The mistake is that presentism and the asymmetry view are both consistent with determinism, and so there is no more reason to think the future is actual in a deterministic universe than in a nondeterministic one. The result is that incompatibilists have the same problems with pre-punishment as compatibilists do.

15. See New, *supra* note 1, at 36–37.

person punished will go on to commit the crime. Subtract doubt of this kind, thinks New, and we subtract concern about pre-punishment also. However, what has just been argued is that there are lingering epistemological sources of concern about pre-punishment *even in the face of absolute certainty that the crime will occur*. If certainty that the crime will occur is certainty that something that is not actual, namely, a future event, will become actual, then punishment for it now is still in violation of the Actuality Principle; the punishment even in such a case is issued for a crime that is not part of the actual world. Because, for all we know, future events are not actual, even if they are certain to occur, concerns about pre-punishment have epistemological sources. But those sources are deep and much harder to overcome than are doubts about what a person will do in the future. In fact, they are so deep that it may not be possible to overcome them even in principle. And even if knowledge in this domain is possible, it is not possible to expect it of those who are busy with the work of placing blame and divvying out punishment. Surely we cannot expect judges or juries, friends, parents, or foreign policymakers, for instance, to jump without either a metaphysical net or a commonsensical one, as it were. For all any of them knows, the asymmetry view is true. Insofar as our concern is with issuing justified punishment, we need to be modest about our state of metaphysical knowledge to the degree that we go beyond common sense. Such modesty mandates against pre-punishment not because punishment in advance is necessarily wrong but because punishment of nonactual crime *definitely* is and for all we know pre-punishment is an instance of it.

The argument just offered depends crucially on the truth of the Actuality Principle. Why should we think it is true? It is possible that there are multiple grounds for accepting the principle. To see one way of understanding its basis, however, consider the following remark from Samuel Pufendorf, the seventeenth-century German moral philosopher:

[A]n obligation differs in a special way from coercion, in that, while both ultimately point out some object of terror, the latter only shakes the will with an external force, and impels it to choose some undesired object only by the sense of an impending evil; while an obligation in addition forces a man to acknowledge of himself that the evil, which has been pointed out to the person who deviates from an announced rule, falls upon him justly, since he might of himself have avoided it, had he followed that rule.¹⁶

16. 2 PUFENDORF, *supra* note 4, at 91. This remark is quoted in part in STEPHEN DARWALL, *THE SECOND-PERSON STANDPOINT: MORALITY, RESPECT, AND ACCOUNTABILITY* 23 (2006). Darwall draws a lesson from the remark very similar to that drawn here.

Pufendorf is here distinguishing between two things: (1) failing to comply with a rule or a demand and thereby suffering harm *because another exercises coercive power*, and (2) failing to comply with a rule or demand and thereby suffering harm *because one has violated an obligation*. The second is the case of justified punishment. The identifying feature of suffering justified punishment for the violation of an obligation, according to Pufendorf, is this: the recipient of the harm in the case of punishment is equipped to recognize the justice of the harm that is received. Following Stephen Darwall, let us label this idea as follows:

Pufendorf's Point: a punishment of person *D* for crime *C* is just only if *D* is equipped to recognize it to be just.¹⁷

Being equipped to recognize the justice of one's punishment, on the one hand, and actually recognizing it, on the other, are importantly different. Pufendorf's Point does not rule out the justice of punishing those who, wrongly, believe their punishments to be unjust. Many of them are perfectly equipped to recognize their mistake.

Still, there is a hard question as to what equipment one must have, exactly, in order to be in a position to recognize the justice of the punishment one receives. But for our purposes what is important is only this much: one must be in position to recognize oneself as identical to the person who committed the crime for which punishment is issued and must do so in a way that involves *caring* about the fact that the act was performed by oneself. To be justly punished one must be in a position to *own* the act as one's own. This involves recognizing identity between the agent of the act and oneself; but it involves more than this. It also involves recognizing that thanks to the identity of oneself with the agent of the act, the act belongs on one's moral ledger; the act is among the things that are rightly appealed to in any moral assessment of oneself, whether an assessment done by another or an assessment of oneself performed by one's self. Self-attribution, in the sense in which the term will be used here, is essentially a moralized notion. We self-attribute

17. Elsewhere, I have argued that this idea is crucial to Locke's theory of personal identity according to which a punished person is identical to the person who performed the bad act only if that person could be conscious of that bad act at the time of punishment. See GIDEON YAFFE, *LIBERTY WORTH THE NAME: LOCKE ON FREE AGENCY* 119–39 (2000); Gideon Yaffe, *Locke on Ideas of Identity and Diversity*, in *THE CAMBRIDGE COMPANION TO LOCKE'S ESSAY CONCERNING HUMAN UNDERSTANDING* 192, 192–230 (Lex Newman ed., 2007).

acts for which we are justly held responsible but do not self-attribute, in the sense in which the term is to be used here, things that we believe to be true of our past selves but that we take to be morally irrelevant. People self-attribute lies they told themselves yesterday, but they do not self-attribute the hair colors of those who told those lies. They know that it was that person whose hair was that color, but self-attribution involves the additional recognition of the self-attributed fact as morally salient, which the hair color is not.

Notice that if no one is equipped to self-attribute the conduct of a nonactual person, then the Actuality Principle follows from Pufendorf's Point. If to self-attribute the crime, the person who commits the crime must be actual, then it follows from Pufendorf's Point that punishment of an actual person is justified only if the person who committed the crime for which punishment is issued is also actual.

So are we ever equipped to self-attribute the conduct of nonactual people? Part of self-attribution involves belief in identity, and it can seem obvious that we are identical to some nonactual persons. Had McCain won the election, *he*, McCain, would have been President. That is, the actual McCain, who is not President, is identical to, or the same person as, a possible, nonactual person who is: namely, the one living in the possible world in which the election went the other way.¹⁸ But McCain is not in a position to self-attribute the quality of being President. It would be newsworthy, to say the least, to read that McCain asserts that he is the same person as one who is President, or that he is proud of having won the election. To read such a thing would be to discover that McCain is delusional. Why? After all, McCain is indeed identical to the nonactual President McCain, the one living in a nonactual possible world. So why is his assertion false? The answer is that in many contexts we take identity talk to be expressive of attitudes of self-attribution. So that when McCain asserts himself to be identical to someone who is President, he is asserting something true, provided that the range of people he is discussing includes nonactual people, but in the context, we assume that he is expressing self-attribution of the qualities of the person who is President. Because the nonactual person's accomplishments

18. One of the greatest metaphysicians of the twentieth century, namely, David Lewis, denied that McCain is identical to the nonactual McCain who won the election. Instead, thought Lewis, the McCain who won the election and the actual McCain are "counterparts." The counterpart relation is related to, but different from, the relation of identity. See generally DAVID LEWIS, ON THE PLURALITY OF WORLDS (1986). Lewis's position helps, rather than hinders, the point sought here because it is clear that when it comes to desert and self-attribution what matters is the relation that actual sufferers of punishment bear to the actual people who committed the crimes for which punishment is issued. However, Lewis's position is not required by the view offered here, and so, in the main text it is assumed that we are identical to our counterfactual selves.

cannot be placed on the actual McCain's moral ledger, he is reasonably taken to be delusional in his attempt to self-attribute the attributes of the nonactual McCain.

That self-attribution of a person's conduct is limited to *actual* people is not some accident of certain linguistic contexts. For a huge variety of purposes, all we care about is whether the person before us, including oneself, is actually, and not merely counterfactually, identical to a person with some important property. It is the actual President, and not the counterfactual one, who is vested, in the actual world, with the power to veto bills and to command the armed forces. It is the actual injurer, and not a merely counterfactual injurer, who is to pay for the hospital bills. It is the actual winner of the race, and not the counterfactual winner, who is owed the gold medal. In all these cases, justified self-directed feelings of guilt, shame, and pride, to name a few, depend not just on identity with a person but also on the actuality of the person to whom the one having those feelings is identical. It is contexts of these kinds, contexts in which it is not just identity that concerns us but identity with actual persons, that are implicated in Pufendorf's Point. We cannot feel guilty for bad things that we know we are going to do in the future but have not yet done. We can feel apprehensive; we can be concerned; but we cannot feel *guilt* for them. The reason is that guilt requires self-attribution of wrongful conduct, and the nonactuality of such conduct undermines the needed self-attribution. Self-attribution of wrongful conduct of a sort that places one in a position to recognize the justice of punishment for that conduct requires the recognition that one is identical to the person who engaged in that conduct *and* that that person is actual. Without actuality we lack the special kind of care or concern with the conduct of the person with whom we take ourselves to be identical that is involved in self-attribution. The Actuality Principle follows.

Notice that one might deny Pufendorf's Point for reasons that fail to impugn the just-offered derivation from it of the Actuality Principle. More is required to be equipped to recognize the justice of a punishment than the capacity to self-attribute the act for which punishment is issued. Among other things surely, one must also recognize that the particular punishment issued is a fitting response to the particular crime for which it is issued, and plausibly one must recognize that the agent issuing the punishment has proper standing to do so, and more besides. Additional equipment would be required to recognize all of these things. It is possible that punishment can be justified even if a person lacks such additional

pieces of equipment, whatever they happen to be. If that is right, then Pufendorf's Point fails. But even if that is so, the point *sought here* stands, for surely punishment is not justifiably issued to a person who is genuinely incapable of self-attributing the wrongful act. From the point of view of somebody lacking that kind of basic equipment, the punishment is no different from punishment for something done by some other person entirely.

The only truly deep obstacle to self-attributing the acts of a future person is the possibility that future persons are not actual. *D* intends to break a promise tomorrow. He has no reason to doubt that he will survive until tomorrow and no reason whatsoever to believe that he will change his mind before the time of action arrives. Nor does he have any reason to believe that somehow, when the time comes, he will inexplicably fail to execute while maintaining his intention. Everything he knows points to his breaking his promise tomorrow. He can truly say without the least hesitation, "I will break a promise tomorrow!" Add that, in fact, it is all true: tomorrow he will indeed break a promise. But is he equipped today in the sense that matters to Pufendorf's Point to self-attribute the acts of the promise-breaker? He is if the future promise-breaker is actual; but he is not, if not. Even if *D* is a committed four-dimensionalist, and so believes himself to be identical to an actual future promise-breaker, he is not equipped in the way that matters *should the asymmetry view be true* despite what he thinks. Recognition is a success-notion: one recognizes *P* only if *P* is true. And so one recognizes oneself to be identical to an actual future person only if one is identical to that person and only if that person is actual. In the absence of the actuality of the future, no one is equipped to recognize oneself to be identical to an actual future person, for there is none. And so in the absence of the actuality of the future, no one is equipped to self-attribute any future conduct. This should not seem surprising or peculiar. For the same reasons, sometimes the only obstacle to knowing something is that it is false. For this reason, none of us is equipped to know falsehoods, although we are unfortunately well equipped to believe them.

As suggested in passing above, it is possible that the Actuality Principle is derivable from more than one appealing principle and not just from Pufendorf's Point. It is even possible that it is a basic axiom that is not in need of derivation from any source. As we will see in Part III, however, whatever its source, the Actuality Principle helps us to understand why a certain class of preventions is justified only if the crimes they avert were imminent.

III. PREVENTION AND THE IMMINENCE OF THE CRIME

As is evident from the quotation from Pufendorf offered in Part I, he intended his point to be specific to punishment and blame. He took the distinctive mark of suffering in response to a failure to meet an obligation, in contrast to suffering in response to failure to respond to a coercive demand, to be that the former, if justified, was inflicted only on one equipped to recognize the justice of one's suffering. So the Actuality Principle, if it has its basis in Pufendorf's Point, is also specific to punishment. We could accept Pufendorf's Point while holding that, in contrast to punishments, justified uses of coercive force, like preventions, do not require that the person harmed in response to noncompliance be equipped to recognize the justice of the harm suffered. And so we need not hold that the Actuality Principle applies to preventions. And so we need not hold further that, for the reasons given in Part I, preventions are no more justified than are pre-punishments. It is possible that although pre-punishments are impugned by the Actuality Principle for the reasons discussed in Part I, preventions are not. In fact, this is right, as far as it goes. However, to end there would be to miss a point of importance: at least one class of justified preventions—bearing, as we will see, a similarity to justified punishments—would be significantly less amenable to justification if it failed to meet a condition that bears a similarity to the Actuality Principle. For this special class of preventions, to be described momentarily, a requirement that the crime be imminent plays a similar theoretical role to the role played by the Actuality Principle with respect to punishment.

Start by distinguishing justified preventions into two kinds. Here, assume that *H*, a harm or deprivation of liberty, is inflicted on *D* in order to prevent *D* from committing *C*, a wrongful or criminal act. And assume that the infliction of *H* for this purpose is justified. Now consider the following:

The Desert Claim: suffering *H* would be deserved for the commission of *C*.

Even though we are restricting our gaze to justified preventions, sometimes the Desert Claim is true and sometimes it is false.¹⁹ In one class of

19. I set aside consideration of cases in which the Desert Claim is false because *H* is less than what is deserved for *C*. *D*'s lethal attack can be prevented by knocking *D* unconscious. Killing deserves much more harm than being knocked unconscious, but it is clearly justified to knock *D* unconscious in order to prevent *D* from killing someone.

cases, the Desert Claim is false because what is deserved for *C* is only slightly less than *H*. Say five years in prison is deserved for a violent robbery. *D* is held for five years and a day and thereby prevented from committing such a robbery. If there is an objection to this act of prevention, and there might be, the objection is not that the harm was undeserved for the act prevented, although it was by hypothesis. That seems a trivial consideration to bring up, given that a violent robbery was prevented. In yet another class of cases in which prevention is at least arguably justified, what needs to be done in order to prevent the crime is so bad as to be undeserved for any crime, no matter how bad. Those who think that it is justified to torture someone in order to prevent a terrorist act by that person might consistently hold that torture is never deserved. On their view, it is justified to torture despite the fact that the terrorist act thereby prevented would not deserve to be punished with torture. Such a position is consistent with, although not entailed by, everything said here.

However, the Desert Claim is often true. *H* is often exactly what would be deserved for the commission of *C*; what is inflicted in order to prevent the crime is often just what the crime would deserve if completed. Arguably, there are cases in which the Desert Claim is true, but the act of prevention is justified without appeal to it. For instance, assume for a moment that child molesters are underpunished; their crimes deserve longer sentences than they typically get. A child molester who is detained after serving a sentence on the grounds that he or she is an incorrigible predator might thereby be detained for as long as would be deserved for the crimes that are thereby prevented. But what justifies detaining the molester, if anything does, is not the facts about what is deserved for child molestation but instead the harms to children that are thereby averted. Arguably, that is, just as the facts about desert are sometimes swamped by consequential considerations in the justification of preventions when the Desert Claim is false, they are sometimes swamped when it is true.

However, sometimes *part of what justifies the act of prevention is the fact that the Desert Claim is true*. Someone who justifies killing a lethal aggressor by saying, "I just gave him what he deserved; he was going to kill me!," is citing the Desert Claim as a defense. So understood, the person is saying that what made killing an appropriate response was that it was deserved for that which the victim would have otherwise done. Let us use the term *Deserved Preventions* to refer to those justified

These cases are probably to be treated very much like those in which the Desert Claim is true.

preventions in which (1) the Desert Claim is true and (2) part of what justifies the prevention is that the Desert Claim is true.

The Deserved Preventions are a special class of exercises of coercive power. They are not instances of giving people what they deserve. The person who is harmed does not deserve the harm because to deserve it, the crime for which it is issued would have to be actual, which, by hypothesis, it is not; it is prevented by the infliction of harm. But the Deserved Preventions are importantly constrained and guided by considerations of desert. In this way they bear a greater similarity to punishments than do other exercises of coercive power.

Forfeiture theories of self-defense—theories according to which it is justified to kill a lethal aggressor because through the aggressor's conduct and intention, he or she has forfeited the right not to be killed by others²⁰—take lethal self-defense cases to be Deserved Preventions. There is a fit, under such views, between that which the aggressor forfeits and that which the aggressor receives; this is the same fit as that between wrongdoing and deserved punishment. Were the aggressor to receive worse than the aggressor forfeited protection from, then the act of prevention, the self-defensive act, would not be justified on such views. So the fact that the harm the aggressor suffers is deserved for the act the aggressor would otherwise have performed—the fact that the Desert Claim is true—plays an important role, under such views, in the justification of the act of prevention. The aggressor only forfeits protection from that which it would be deserved to receive for the aggressive act.

The Deserved Preventions are, in an important sense, a peculiar class. We can understand why the fact that *D* deserves, in fact, to suffer harm *H* helps to justify inflicting it on him. But how can the fact that *D would* deserve harm *H*—*were things different*—justify *actually* inflicting *D* with *H*? If the Actuality Principle were false, it would be easy to answer this question: *D* is identical to the nonactual person who commits crime *C*—in the possible world in which *D* is not prevented from doing so. But this will not do; the actuality of the crime is needed to justify the actual infliction of the punishment for it; the Actuality Principle is true. But still, the Desert Claim does play a role in justifying some preventions. How does it do so? Why are counterfactual considerations about desert relevant here?

20. E.g., Judith Jarvis Thomson, *Self-Defense*, 20 PHIL. & PUB. AFF. 283 (1991).

The reason that they are relevant is because the justification of prevention through appeal to the Desert Claim proceeds by comparing two worlds *in which desert is maximized*: a world in which the act of prevention takes place, and a world in which it does not. If we maximize desert in both worlds, then in the nonprevention world, *D* commits *C* and receives *H*. *D* receives *H* in that world because *H* is deserved for *C* and we are imagining that desert is maximized in both worlds, consistent with one being a prevention world and the other a nonprevention world. That is the import of pointing to the Desert Claim in justification of the act of prevention. In the prevention world, *D* receives *H* and does not perform *C*; in that world, *D* receives something undeserved, for, at best, *D* will deserve to receive *H* only if *D* performs *C*, which by hypothesis *D* does not because *D* is prevented from doing so. If we now compare the value of the two worlds, we find that they do not differ in one important way: *D* receives *H* in both. But they do differ in two ways: first, in the prevention world, in contrast to the nonprevention world, *D* receives a harm he does not deserve; and second, in the nonprevention world, in contrast to the prevention world, he performs *C*. The first of these two differences counts against the value of the prevention world; the second against the value of the nonprevention world. If the calculation is to come out in favor of the prevention world—if that is to be the better world than the alternative world, despite its involving some undeserved harm—then the degree to which *H* is undeserved in the prevention world should be minimized. It would be minimized if, in that world, *D* gets as close as possible to performing *C* while still being prevented from doing so. The closer *D* gets to committing *C*, the closer the prevention world is to a world in which *C* is actualized. And so the closer that world is to one in which *D* deserves the harm that he is issued in order to prevent him from committing *C*. What must the prevention world be like in order for this to be the case? Answer: it must be one in which *D*'s performance of *C* is imminent.

Imminence is not actuality. If the future is not actual, then neither is the imminent future. But worlds in which a crime is imminent, but does not take place, are closer to worlds in which it does take place than are worlds in which the crime is not imminent. Closer in what respect? In the respect that matters—there will be more on this momentarily. Justification of prevention through the Desert Claim, then, proceeds in a way that motivates an imminence requirement: among the worlds in which the crime is prevented, the world in which it was imminent is closer than any other in the set to the world in which the crime is not prevented and is given what it deserves. Thus to realize the world in which the crime is imminent is to be closer than any alternative to a world in which the recipient of the harm is given what the recipient

deserves. It is, still, to fall short of giving what is deserved, but it is to fall short of it in service of the goods that are realized by preventing the crime.

It is worth putting this important point in somewhat different language. The idea is this: we are on firm ground when we give individuals what they deserve, even if we harm them in doing so. However, deserved harms are governed by the Actuality Principle: we give one a harm one deserves only if the crime for which the individual deserves it is actual. Given this, it follows that successful preventions are never, by definition, given to someone who deserves them: they always serve to prevent crimes from being actualized. We therefore need to know why the fact that the harms involved in them would be deserved were the crime committed is of any relevance at all to their justification. The answer is that it directs us to compare the actual world in which the crime is prevented by harming the person who would otherwise commit it with a world in which it is committed *and is given what it deserves*, namely, just what it is given in the actual world. The world in which the crime is given its just desert is relevant because one who weighs desert heavily in one's calculations about what to do cares about such a world; from the standpoint of desert, which is only one value among many but a value still, that world is ideal. The question is, How far short of the ideal desert world is one falling in acting to prevent the crime? Justifying one's conduct, conduct that causes the world to fall short of the ideal of desert, requires showing that the gap between the ideal from the point of view of desert and the actual world is minimized consistent with other values, in particular those that are furthered by preventing the crime. Imminence allows for the minimization of that gap because the world in which the crime is prevented when imminent is closer to, more similar to, the world in which the crime occurs and is given what it deserves than is any other world in which the crime is prevented. Because we care about desert, we care about actuality; because we care about actuality, we want to act so as to treat people contrary to what they deserve in such a way as to minimize our departures from worlds in which they are given what they deserve; because we want to minimize such departures, we care about imminence.

The justification of imminence requirements that has just been offered shows how they are distinct from necessity requirements. One can know for certain that the crime will be committed tomorrow if one does not act today. But to act today is to realize a world that is further from one in

which the person harmed gets what is deserved than the world that results from acting tomorrow—when the crime is imminent—would be. If all that matters is preventing the harm, then this concern should not move us. The harm is prevented whether we act today or tomorrow. But if giving people what they deserve also matters, then we should wait until the crime is imminent.

It would be a mistake to conclude from this discussion that we ought to convict of murder the woman who kills her abusive spouse or boyfriend while he sleeps in order to prevent him from killing her when he wakes.²¹ We may decide that, especially given his history of awful behavior, the killing is justified; perhaps he had it coming, even though his attack was not imminent. But if we reach that conclusion, we ought to do so while recognizing that there is an important sense in which he has been wronged: he has been given something that he did not deserve—he did not deserve to be killed *for killing her* because that crime was not actual. And it is quite possible that that wrong could have been lessened by waiting until his crime was imminent. Perhaps this is too weak a hook to hang a murder conviction on, however; if it is, then we ought to acquit, or at least mitigate to some lesser grade of homicide, or even to a grievous assault, when imminence is absent. The point is that whenever a person harms another for reasons of prevention, there is a trade-off of values: whatever values are furthered by preventing the crime are to be weighed against the disvalue of giving someone a harm or deprivation of liberty that the individual does not deserve. We can minimize the gap between these values by waiting to prevent until the crime is imminent. But it is possible that the calculation will come out in favor of prevention even in the absence of imminence, and in such cases prevention is still justified.

There is an objection to this line of thought that must be addressed. The argument proposed assumes that if World 1 resembles a world in which *D* gets what is deserved more than World 2 resembles it, then the person who values desert ought to prefer World 1 to World 2. The objection is that *with respect to desert*, World 1 and World 2 might be equally far away from the world in which *D* gets what is deserved. A red shirt resembles red pants more than a green shirt resembles red pants, but for the person in need of *pants*, that is not a reason to choose the red shirt. Similarly, given that the infliction of *H* is undeserved whether or not *C* was imminent, in what sense are we getting closer to giving people what they deserve when waiting to inflict *H* until *C* is imminent? The best answer to this objection may not be, in the end,

21. See, e.g., *State v. Leidholm*, 334 N.W.2d 811, 813–14, 818–20 (N.D. 1983).

entirely satisfactory. Still, it is worth proposing. The answer adverts to the justification of the Actuality Principle offered in Part II. There it was suggested that the Actuality Principle is true because self-attribution of wrongdoing of the sort that underlies the recognition of a punishment to oneself as deserved requires the actuality of the self-attributed act. But introspective attention to the self-directed attitudes that become appropriate when a bit of wrongdoing is imminent, but not actual, demonstrates an important similarity between such attitudes and those that are appropriate when wrongdoing is actual. There is something *like*, although different from, guilt that is appropriately felt when wrongdoing is imminent. There is something more *like*, although different from, self-attribution possible when wrongdoing is imminent than when it is not. These are empirical, psychological claims, and nothing more than the results of one person's introspection is being offered in their defense. But if these claims are true, then there is an important disanalogy between aiming at desert and so choosing a world in which the objectionable conduct is imminent, on the one hand, and aiming at red pants and so choosing a world in which one receives a red shirt, on the other. Imminence is like actuality in a way that matters—it underwrites an attitude *like* self-attribution, which is the attitude of importance that is underwritten by actuality—while shirts are entirely unlike pants in the way that matters, regardless of their color. In short, through appeal to imminence we can clothe naked exertions of coercive power in which undeserved harms are inflicted.

Although, as just indicated, the idea that something closer to self-attribution is possible with respect to imminent future action than with respect to future action that is not imminent is an empirical claim for which no adequate evidence is being offered, reflection on a related proposition for which empirical evidence is available can buttress our confidence in it. In recent years, a substantial amount of empirical evidence has been produced indicating that animals, including human beings, often temporally discount future goods hyperbolically.²² By giving subjects choices between present rewards—\$10 now, for instance—and greater future rewards—\$100 to be delivered a year from now, for instance—it is possible to plot the degree to which subjects discount the value of future goods in their decisionmaking as a function of their degree of

22. See, e.g., Shane Frederick, George Loewenstein & Ted O'Donoghue, *Time Discounting and Time Preference: A Critical Review*, 40 J. ECON. LITERATURE 351, 360 (2002).

futurity. A subject, for instance, willing to take \$10 now over \$100 a year from now but unwilling to take \$9.99 today discounts the future \$100 by 90%. If we move the clock forward 364 days and find that the subject is now willing to take, say, \$11 today over \$100 a day from now but not \$10.99, then the subject's discount of the future reward only decreases by 1% over a 364-day period. Assuming that a day later, when the choice is between \$100 now and something less now, the subject will not take anything less than \$100, the subject must radically increase valuation of the reward over the final day before its acquisition. When subjects' discount curves have been plotted and equations have been used to describe them, it has been found that, in a wide variety of circumstances, subjects discount future goods hyperbolically: they discount them a very great deal until the final unit of time before their acquisition, at which point their degree of discounting decreases very rapidly. The curve plotting the value that the subject assigns to the \$100 reward, for instance, relative to time, increases very slightly from \$10 to \$11 over the first 364 days and then jumps rapidly up to \$100 over the final day. Put in the language of relevance to the project here: imminent future goods are granted much more value than future goods that are not imminent. This is a robust empirical finding.²³

However we characterize, precisely, the psychological attitude towards future goods that has been found to increase hyperbolically with the passage of time, this much is clear: it is in some important way intertwined with the attitude of self-attribution. It is the *anticipation* of future goods that alters in this hyperbolic way. We do not discount goods for others, whose welfare matters not at all to us, in the same way as we do goods for ourselves. It is not an attitude of detachment towards future goods that is being measured but instead an attitude that involves the recognition that it is *oneself* who will be enjoying the future good that is discounted hyperbolically. It is some measure of the degree to which we care about a future good in light of the fact that it is ourselves who will be receiving it. The claim that concerns us here—that our attitude towards future conduct is closer to the kind that is required for desert of punishment for that conduct, closer to self-attribution, when the future conduct is imminent—is closely parallel to the empirically supported claim that the attitude of anticipation of a future reward is much more like the attitude involved in the actual enjoyment of that reward when the reward is imminent than when it is not. To be clear, what is being proposed is not that empirical findings of a stronger attraction towards imminent future goods in comparison with distant ones are, full stop, supportive of the

23. For a marvelous and comprehensive review of the literature on this topic, see generally *id.*

claim that there is something more like self-attribution possible when future action is imminent than when it is distant. Rather, what is being proposed is the more modest claim that there is a suggestive connection between these two things, suggestive enough to warrant greater confidence in the claim that is required to support the argument offered here than would be warranted merely through a flat appeal to introspection.

A thought experiment can be used to illustrate the point. Imagine that you have such a sensitive conscience that whenever you are performing an act, your conscience rewards or punishes you in precisely the way that you deserve to be rewarded or punished while you are performing the act. While doing wrong, you are punished, right then, with the pangs of conscience that are precisely deserved for that wrongdoing. And assume that you discount future goods and harms hyperbolically: temporally distant future harms matter hardly at all to you now, but temporally imminent harms matter to you a great deal, and nonimminent harms matter to you very little. In such a case, there would be an obvious sense in which your attitude towards future wrongdoing on your part would be more like the attitude of self-attribution needed for desert of punishment when the wrongdoing is imminent than when it is not. In the last unit of time before your bad behavior you would be feeling towards your future self much like you will be feeling towards yourself when you actually engage in the bad behavior; you will be anticipating the future pangs of conscience in a way that approximates what it will actually be like to suffer them much more closely than you were before the bad behavior is imminent. If the kind of self-attribution of wrongdoing that is needed for desert of punishment is like the kind of self-attribution that is involved in the pangs of conscience, then it is possible to see how the facts about our tendency to hyperbolically discount are related to the facts about our capacity to self-attribute—in the way that matters—imminent wrongdoing in contrast to temporally distant wrongdoing.

It is important to highlight an aspect of the view proposed here that might not be entirely obvious. The view proposed involves a novel understanding of the importance of possibilities for change of mind on the part of a person against whom we act so as to prevent us from behaving badly. One might think that what is crucial about the imminence of bad behavior is that, in its absence, there is still an opportunity for aggressors to change their minds. Add to this the thought that, when individuals still have the opportunity to prevent *themselves* from acting badly, to prevent them from doing so fails to respect their autonomy, and you can be led

to the conclusion that imminence is required by our respect for autonomy. A first point to note about such a view is that it hardly respects the autonomy of individuals who are committed to *not* changing their mind to refuse to act against them until their opportunity to do so has passed. To take such an attitude is to fail to respect their present resolve to not changing their minds. It is patronizing to say, “You might feel differently when you are older,” and it is no less patronizing, perhaps even more so, to say that when the envisioned future is only minutes or hours away. It is in part because we *respect* the autonomy of those on whom we rely that we rely on them to do the things in the future that they are committed, now, to doing.

But still, there is an important difference between acting against one who still has opportunities to change one’s mind, on the one hand, and acting against one who does not, on the other. Under the view presented here, the difference concerns the capacity of each of these two people to self-attribute the future bad conduct. Neither can self-attribute the future conduct in the way that would be required, under Pufendorf’s Point, for just punishment for that conduct. But the latter, who sees no opportunity to change his or her mind in the future, is in the position to have feelings closer to feelings of guilt about that future conduct than are people who recognize that it is still up to them whether they act as they are resolved to act. That is, recognizing that you still might change your mind serves as an additional obstacle to associating yourself with future conduct, even if you are resolved not to change your mind. This, in turn, makes one an even less fit object of preventive harm—even harm that would be deserved for the future bad conduct—than one would be were there no future opportunity to change one’s mind. The result is that when there are no opportunities to change one’s mind, the world is closer, in the ways that matter to Deserved Preventions, to a world in which the inflicted harm is, indeed, deserved. The possibility of changing one’s mind matters, and weakens the case for preventive harm, because of its relevance to one’s capacity to self-attribute future conduct and not because of its alleged link to respect for autonomy.

IV. CONCLUSION

In the introduction to this Article, a pair of puzzles was identified. First: why is it that preventions can be justified when pre-punishments cannot? Second: why is it that successful preventions are morally better than unsuccessful preventions, while pre-punishments that avert the crime are morally worse than pre-punishments that do not?

The answer to the first puzzle is that unlike preventions, punishments are justified only if deserved. Because desert requires actuality, and

because for all we know future events are not actual, it follows that pre-punishments are not justified. Preventions, by contrast, always involve harming people in a way that they do not deserve. They are justified—or, in any event, the Deserved Preventions are justified—to the degree to which they maximize their similarity to the issuance of deserved harms—something that they do when the crime prevented was imminent.

We also have tools to solve the second puzzle. In both the prevention case and the pre-punishment case, there are three worlds to consider: World 1, crime followed by deserved harm; World 2, harm followed by crime; and World 3, harm followed by no crime. The puzzle is that when we are considering punishment, World 2 seems less bad than World 3, although when we are considering prevention, World 3 seems less bad than World 2. The explanation for this derives from the kind of considerations that are relevant for justifying punishment in contrast to those that are relevant for justifying prevention. In particular, averting the very crime for which the punishment is issued—as opposed to averting future crime by that very offender or others—plays no role in the justification of punishment. Because it plays no role, it cannot count in favor of the value of World 3 when the issuance of the harm is being considered as a punishment, but it can count in favor of that world when the issuance of the harm is being considered as a prevention. Put another way, World 2 is similar to World 1 in a way that matters for punishment; there are both harm and crime and so there is fit between them. But World 3 is dissimilar to the other two worlds in the way that matters to prevention: the values associated with preventing the crime are furthered in that world but not in either world in which the crime occurs.

The lingering question is why this fact about World 3 can play no role in the justification of punishment. The reason is that to allow it to play such a role would be to run afoul of the Actuality Principle. To allow the prevention of the very crime for which a punishment is issued to play a role in the justification of that punishment is to allow a punishment to be justified in violation of the Actuality Principle. This is, in turn, if the argument of Part II concerning Pufendorf's Point is granted, to allow that a punishment can be justified even when issued to a person incapable of self-attributing the crime for which it is issued. These are untenable results that we do not find when we turn to the case of prevention. There we start from the assumption that we are engaging in a practice that involves inflicting undeserved harms; we are exercising coercive power. The question is only how to minimize the damage for the sake of other

goods. When that is the question on the table, principles like the Actuality Principle or Pufendorf's Point play a secondary role. This is not to say that such principles play no role: as we saw in Part III, it is thanks to them that we can see why an act of prevention is better when the crime is imminent than when it is not, even if the crime was inevitable either way in the absence of preventive measures being taken. But still, such principles do not place direct constraints on the justification of prevention as they do on the justification of punishment.

There is a tradition of claiming that one is justified in inflicting undeserved harm on another merely because that infliction is not actually *punishment* but something else and only punishment must be deserved to be justified. The worry about such claims is that they place normative weight on what is, at best, a semantic distinction. However, if the argument of this Article succeeds, it demonstrates that, at least in the case of prevention, noting how the act differs from punishment serves to immunize it from criticisms that would be rightly leveraged against a punishment that involved infliction of the very same harm. The justified preventer is striving to minimize the damage that the preventer does to what in Pufendorf's time was called "the order of grace"—the balance of wrongs with deserved punishments—while at the same time striving to further the values that are damaged through crime. This is different from punishing because it involves a different and less direct concern with desert. But insofar as it involves any concern with desert, the justification of prevention encounters the need to accommodate the fact that harm or deprivation of liberty is deserved only if issued for an *actual* crime. As has been suggested here, in the case of prevention, some steps can be made to make this accommodation through waiting to inflict harm until the crime is imminent.