

THE CONSEQUENTIALIST PROBLEM WITH PREPUNISHMENT

FORTHCOMING IN *THOUGHT*

ABSTRACT. This paper targets a nearly universal assumption in the philosophical literature: that prepunishment is unproblematic for consequentialists. Prepunishment threats do not deter, as deterrence is traditionally conceived. In fact, a pure prepunishment legal system would tend to increase the criminal disposition of the grudgingly compliant. This is a serious problem since, from many perspectives, but especially from a consequentialist one, a primary purpose of punishment is deterrence. I analyze the decision theory behind pre- and postpunishments, which helps clarify both what deterrence is and how it operates in consequentialist justifications of punishment. I end by sketching a road map for the future of prepunishment as artificial intelligence and other technological advances generate increasing possibilities for its use.

If the police could know that a person will commit a crime tomorrow, then could they reasonably decide to “prepunish” the predicted offender today? This was the question first asked by Christopher New [1992], who argued that such a practice, if made feasible, would be morally permissible. The intuition against prepunishment, New claimed, results from “attaching improper moral significance to an insignificant temporal fact” [37]. Subsequently, many philosophers have disagreed with New on nonconsequentialist grounds.¹

Consequentialists think that the practice of punishing should aim to produce good consequences. One of the primary ways the threat of punishment produces good consequences is by deterring crime. And the most obvious way that the threat of punishment deters crime is by making people fear that they would be punished in response to their criminal act. Virtually everyone assumes that prepunishment is unproblematic, or dramatically less problematic, for consequentialists. For those who think that prepunishment *is* problematic, this is viewed as a reason to doubt consequentialism.

Underlying the belief that prepunishment is unproblematic for consequentialists is the assumption that the deterrent value of prepunishment is similar to the deterrent value of postpunishment. This starts with New (1992, 38): “In pre- as much as in postpunishment the penalty imposed may deter potential offenders, as also the actual offender, from committing other offences in the future.” In his influential reply, Smilansky (1994, 50) agrees with New at least to that extent. He writes, “We need not concern ourselves here with [consequentialist objections]. I allow prepunishment may be useful. . . . The crucial question is whether someone who believes in nonconsequentialist constraints of justice has the resources to reject [repunishment].”

Subsequent discussions of prepunishment have emphasized its supposed deterrent value. Statman (1997, 129) writes, “Utilitarian benefits that can be gained by punishing after crime (‘postpunishment’) can also be gained by prepunishment, e.g., deterring other potential criminals. . . . It is a puzzle mainly for retributivists.” Petersen (2014, 140) writes:

¹A separate but rich line of inquiry concerns the relationship between prepunishment and free will. This debate started with Smilansky (2007) and has also proven popular.

According to a utilitarian rationale, the purpose of punishment is crime reduction (by means of strategies like rehabilitation, incapacitation, or deterrence). And if, for example, a neuroprediction based on a brain scan of an offender X's future behavior tells us that X is likely to commit new crimes in the future, there is a utilitarian reason to punish X more harshly than if X were predicted not to be dangerous.

Surprisingly, and to the contrary, I will argue in this paper that the question of whether and how prepunishments deter criminals is much less clear than with postpunishments. On one of the simplest ways of understanding deterrence — giving potential criminals reason to fear the consequences of their criminal act — the threat of prepunishment fails to deter entirely. In fact, when prepunishment threats are not checked by postpunishments threats — as in a “pure” prepunishment system — we should expect an *increase* in criminal dispositions (the distinctive effects of pre- and postpunishment are spelled out in Section 1). To avoid this, a society that engages in prepunishment must ensure that its prepunishment abilities never become more reliable than its postpunishment abilities (the reasoning behind such strategies is presented in Section 2).

As we will see, reflecting on the permissibility of prepunishment from a consequentialist perspective helps to reveal previously unnoticed insights concerning the nature of deterrence, and, in turn, how deterrence features in consequentialist justifications of punishment. Consequentialist problems also matter for questions of application, for at least two reasons. First, the practice of prepunishing exists in society today — though the extent of it is debated. Most ambitiously, some philosophers and legal scholars argue that prepunishments are sometimes involved in real-world *criminal convictions*.² More modestly, individuals do sometimes aim to prepunish others on the basis of their beliefs about what the other person will or would do. Second, as rapid advancements in artificial intelligence increase our capacities to prepunish — to engage in predictive policing or preventive detention based on predictions of future crimes (as in parole decisions) — society would benefit from an understanding of how prepunishments succeed or fail on consequentialist grounds.

1. PUNISHMENT-PROBABILITIES, OUTCOME-PREFERENCES, AND SELF-BINDING

Consider a thought experiment. Your company, MegaTech, has laid off its payroll staff and now leaves payroll to the management of individual employees. The plan is simple: at the end of the week, you are given access to the vaults so that you can take your salary. To keep employees from stealing, predictions are made on Monday about whether an employee would steal on Friday. These predictions have been found to be reliable enough to establish beyond a reasonable doubt that an employee would steal. Any predicted thieves are fired.

Now consider two ways the story might continue.

Counterfactual

On Monday you are not fired. Now it's Friday. You're alone in MegaTech's vaults. It occurs to you that you could safely take more than your salary, and no one would notice. You are confident that MegaTech won't fire you, since any firing would have happened on Monday. So, you decide to take some extra money and, of course, you are not fired.

²See Beebee, 2008 ; Greene, 2008; and Zedner, 2007.

Actual

On Monday you receive word that you have been fired. “Why?” you protest. “I wasn’t planning to steal any money! Even if I want the money, I know about the risk of punishment, and it’s not worth it to me!” Your boss replies, “Ah, but you see, the predictors saw that you were going to steal, for consider *Counterfactual*.”

Counterfactual is a description of the closest possible world to *Actual* in which you are not fired. On Monday, the MegaTech predictors ask, “would you steal if we don’t intervene?” Given *Counterfactual*, the answer is “yes.” And so you are fired.

If MegaTech might postpunish you, then you have a self-interested reason on Friday not to steal. However, with only the prepunishment policy in place, there is no such deterrence on Friday. Imagine that you’re an ideal candidate for deterrence. While in the vault on Friday, you explain: “If there were any risk whatsoever that taking this money would cause a punishment, I wouldn’t take it. In other words, I am *maximally deter-able*. But there is no such risk; nothing I do now can change the past.” Thus, MegaTech’s threat of prepunishment would have no deterrent power on Friday whatsoever. Even a very weak system of postpunishment, which detects criminals only a small percentage of the time, would have more deterrent power on Friday.³

The case of MegaTech is inspired by Philip K. Dick’s influential prepunishment story, “The Minority Report.” In “The Minority Report,” the prepunishment policy aims to prevent crimes and it does not require that a person first form an intention to commit their predicted crime. However, some proponents of prepunishment in the philosophical literature have argued that prepunishment is only justified when people do form the relevant intention or the crime is allowed to occur. No matter: these requirements don’t affect the deterrence problem.

Consider a non-preventative variant of MegaTech’s prepunishment threat in which the prepunishment is a half-dozen lashes to be administered on Monday. This too would have no deterrent effect on Friday. Whether or not you receive the lashes on Monday, on Friday you know that you could take more than your salary without causing any (further) punishment. Thus, whether the threat of prepunishment is preventative or non-preventative is irrelevant to its deterrence problem.⁴

³We should be careful to distinguish between two questions: i) would *being prepunished* deter you from stealing? And ii) would MegaTech’s *credible threat to prepunish you for stealing* deter you, or anyone else, from stealing? The important question for our purposes is (ii). The answer to (i) is, of course, obviously no — it is the *threat* of punishment that might deter you and not the actual punishment itself. Nevertheless, we should note that a consequence of actually prepunishing someone is to provide others with evidence that the threat to prepunish is credible. As an anonymous referee points out, actual prepunishments could also provide evidence that *postpunishment* threats are credible (e.g., by increasing confidence in the overall abilities of police). These effects of actual instances of prepunishment would need to be taken into account in a total consequentialist analysis of prepunishment policies in a specific context. However, for the purposes of this paper, I stipulate that the prepunishment threat is known to be credible, so that our focus is on determining whether the credible threat to prepunish deters (our target question), and not on the question of how a justice system might make its threats seem credible.

⁴This paper focuses on preventative prepunishment because consequentialist punishment justifications tend to favor prevention when available, and especially so when a crime inflicts harm (see Milevski, 2014). (Prepunishment implies that we know ahead of time that a person would commit a crime, in which case there is the opportunity to prevent harm). As several authors have conceptualized the issue (e.g., Smilansky (2007, 347), Beebe (2008, 259), and Wringer (2012, 135, fn. 4)), prepunishment can be either preventative

Whether an intention is needed also has no effect on the deterrence problem. Regardless of whether you intended on Monday to steal the money on Friday, on Friday you would know that you can steal without causing punishment. At that point, the threat of prepunishment is irrelevant, because you can't change the past.⁵ Thus, MegaTech's prepunishment threat, whether it requires an intention or not, does not deter anyone on Friday.⁶

If the threat of prepunishment does not deter in the traditional way, what does it do? To answer this question, we need to break up the concept of deterrence into smaller components. On a broad conceptualization, deterrence policies aim to reduce crime through the threat of punishment. This can be done in several ways. On the one hand, the threat of postpunishment deters by influencing a person's (actual or anticipated) assessment of the *probability of punishment conditional on a criminal act, at the time of decision*. This is traditional deterrence: credible threats of postpunishment give people reason to believe that if they perform the act in question, then undesirable outcomes, in which they are punished, are probable.

On the other hand, threats of prepunishment deter by influencing a person's *preferences over outcomes* or by causing them to *self-bind*. If your company adopted MegaTech's prepunishment policy, you should ask yourself honestly: "Would I prefer to steal money if I knew I could get away with it?" If the answer is yes, then you will be fired on Monday. You need to change your preferences (so that you prefer not to steal even in situations where you know you can get away with it), or you need to self-bind yourself so that the criminal act is never an option in your future decision situations. These are the effects of successful prepunishment threats.⁷

or non-preventative. Nevertheless, as an anonymous referee points out, others might insist that the term 'prepunishment' only apply to cases of non-prevention (i.e., to cases where a person is punished for something they *will do* and not something they *would have done* had they not been apprehended). Someone with that view would need to come up with a new term for cases of 'preventative prepunishment,' since those cases differ in important ways from the practices of postpunishment and non-punitive prevention. Regardless, this issue is irrelevant to the claims of this paper, since as I show in the main text, the deterrence problem applies to preventative and non-preventative 'prepunishment' equally.

⁵The intention-required version of the MegaTech case mirrors Kavka's (1983) famous Toxin Puzzle. As Kavka argues, you have a reason to try to get yourself to intend to drink the toxin today; you have no reason to actually drink the toxin tomorrow. The same is true in the intention-required version of the MegaTech case. MegaTech's policy would give you a reason to try to refrain, somehow, from forming the intention to steal on Friday, but it would not give you a reason to refrain from stealing on Friday.

⁶Decision-theory enthusiasts may notice a similarity between cases involving prepunishment and Newcomb's problem. While I do not have space to thoroughly discuss the issue here, I note that non-preventative prepunishment systems create "transparent" Newcomb problems (Gibbard and Harper, 1981, 181–2) and preventative prepunishment systems create "Parfit's Hitchhikers" (Parfit, 1984, 7). Since both evidential and causal decision theorists choose the "two-box" equivalent in transparent Newcomb problem and Parfit's Hitchhiker, neither form of prepunishment system deters them (in the traditional sense). Some proposed decision theories do recommend the "one-box" equivalent in these cases, and thus imply that rational agents should be deterred by prepunishment threats. See Gauthier 1986, Chapter 6; McClennen, 1990; Meacham, 2010; Greene, 2018; Soares and Yudkowsky, 2018; and Levinstein and Soares, 2020.

⁷Notice that deterrence theorists have traditionally *not* referenced these effects in supporting their favored deterrence policies. Thus, thinking through the differing deterrent effects of post- and prepunishment can help clarify the nature of deterrence as it features in justifications of punishment.

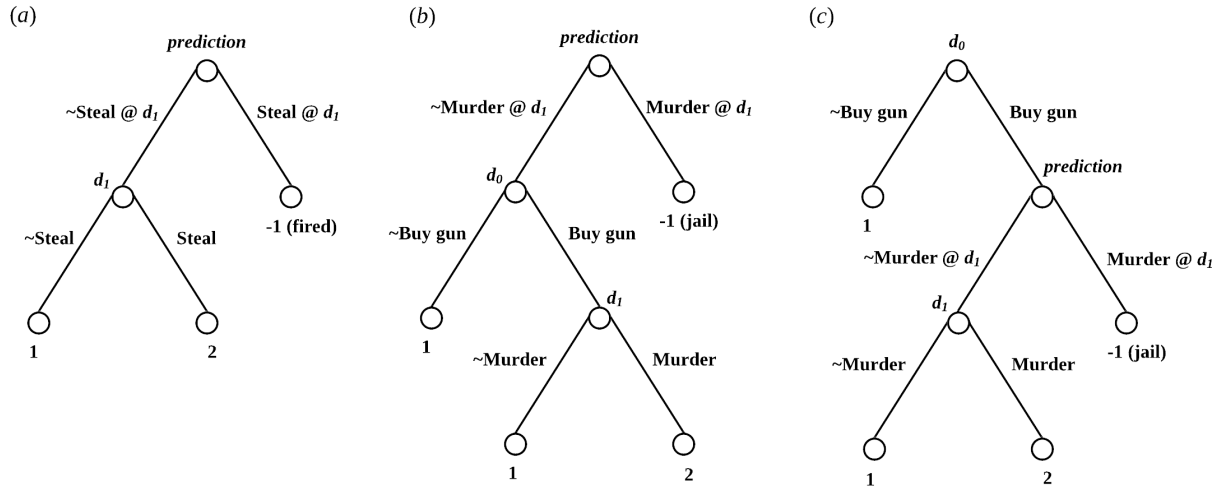


FIGURE 1. Decision Trees for Prepunishment Situations

Using this terminology, we can say that postpunishment threats affect punishment-probabilities, while prepunishment threats affect outcome-preferences and self-binding behavior.⁸ When both kinds of threat occur simultaneously, there is a mix of all three effects.

In the next section, we take a closer look at the deterrent effects of pure prepunishment and mixed pre- and postpunishment legal systems.

2. DETERRENT EFFECTS OF PREPUNISHMENT LEGAL SYSTEMS

Consider three ways a person might respond to the threat of criminal postpunishment. First, *willing compliance*, in which they comply with a law and would do so without the threat. Second, *grudging compliance*, in which they comply with a law but would not do so without the threat. The third way is *noncompliance*.

Whether we threaten prepunishment or postpunishment does not matter to the willingly compliant or noncompliant, but there is a dramatic difference in the response of the grudgingly compliant. The implications depend on whether the prepunishment system is ‘pure’ or ‘mixed.’ Consider each in turn.

2.1. Pure Prepunishment Systems. In a pure prepunishment system, prepunishment has replaced postpunishment completely. A pure prepunishment system might be viewed, as it is in “Minority Report,” as the natural culmination of a legal system as its prepunishment abilities increase. What need is there for postpunishment threats once a system’s prepunishment mechanism captures all potential offenders?

“Minority Report” overlooks the fact that pure prepunishment systems create a serious *hazard* for the grudgingly compliant. This hazard is on display in the original MegaTech case. Imagine that you are grudgingly compliant and consider the decision tree in Figure 1a.

⁸Punishment threats have other potential effects, of course, such as to communicate and inculcate society’s norms. I leave these to the side because the threat of pre- and postpunishment are similar in these effects.

Consider your decision situation at node d_1 : you can either steal money from the vault or take your regular salary. Since you are only grudgingly compliant, and at this point there is no chance of punishment, you would choose to steal. However, before this, MegaTech's predictors predict whether you would steal at d_1 . If they predict that you would, they fire you. Therefore, you never reach d_1 , and instead end up fired.

Granted, the lack of postpunishment makes no difference to the willingly compliant — they're willing to keep within the law even if they could get away it. It also makes no difference to the noncompliant — they're not deterred by the threat of postpunishment in the first place. However, for the grudgingly compliant the difference is crucial: whereas in a postpunishment system the grudgingly compliant decide to keep within the law for fear of the consequences, in a pure prepunishment system they may be punished before they have a chance to decide.

Extrapolating from this, we can expect the result of MegaTech's policy to differ from what its executives intended. Shifting from a postpunishment to a prepunishment policy results in a drastic increase in the criminal disposition of grudgingly compliant employees and results in their mass purge from the company. Something similar is true at the level of society: enacting a pure prepunishment system would tend to drastically increase the number of people sent to jail because it would convict the noncompliant *and* the grudgingly compliant. It would, in effect, result in a mass purge of all but the willingly compliant.

People disagree about the point of criminal punishment: is it to prevent future crimes, give the guilty what they deserve, internalize society's norms, or something else? Whatever the case, the point is not to round up all the people who would commit crimes if they could get away with it. But that is what a pure prepunishment system would do.⁹

Perhaps, however, there are ways for the grudgingly compliant to avoid punishment in a prepunishment system without becoming willingly compliant. If, for example, MegaTech forewarns employees of the new policy before it is enacted, then any grudgingly compliant employees will have a chance to quit, self-bind, or change their preferences before predictions are made.

Consider a related case involving the prepunishment of murder. John, who lives under a pure prepunishment legal system, would murder his ex-wife, Joan, if he could get away with it. However, John would only kill Joan if he had a gun. Could the threat of prepunishment deter John from buying a gun? This depends on when the prediction occurs. If the prediction occurs before John decides whether to buy a gun, and John knows this, then the threat of prepunishment fails to deter him.¹⁰ This is represented by Figure 1b. In this case, the relevant prediction concerns d_1 : John's decision whether to murder Joan once he has a gun. However, the threat of a prior prepunishment does not give John a reason to refrain from

⁹In the world of "Minority Report" there are very few criminal convictions. Dick explains that this is because people do not attempt to commit crimes because they know that any potential future crime would be successfully predicted and prevented. Milevski (2014, 10) calls such a system "strong" prepunishment, and he claims that a strong prepunishment system would be "a perfect system of crime prevention." Pace Dick and Milevski, absent some story about rapid changes to people's outcome-preferences or self-binding behavior, we should expect criminal convictions in a strong pure prepunishment system to skyrocket. Even consequentialists may not view this situation as perfect.

¹⁰Determining how John (i.e., a grudgingly compliant rational agent) would behave under uncertainty about the time of prediction or uncertainty about what has been (previously) predicted involves principles of decision theory on which causal and evidential decision theorists disagree. See Footnote 6 for more on the connections between prepunishment decision situations and decision theory.

murdering Joan at d_1 , and thus if he reached d_1 he would choose to murder her to attain his highest utility outcome (2). By the same inference, the threat of prepunishment also does not give John a reason to refrain from buying a gun at d_0 : nothing John does at that point can change what has already been predicted.

Now consider Figure 1c, in which the prediction occurs only after a gun is bought. John's decision at d_0 is upstream of the prediction that could send him to jail, and thus what he decides can affect what is predicted. If John never buys a gun, then he would never commit the murder. Thus, John would not be prepunished for the murder if he does not buy a gun.

We can think of this strategy as involving a type of self-binding. Choosing not buy a gun at d_0 ensures that John will never be in a position to murder: he will never reach d_1 .

The viability of self-binding strategies in response to the threat of prepunishment motivates the adoption of a “substantial-step” requirement for prepunishments, which mirrors a similar requirement for the prosecution of attempt offenses.¹¹ If a person cannot be convicted of a prepunishment crime until a substantial step is undertaken (such as buying a gun), then would-be offenders can be deterred from taking a substantial step. If, however, prepunishment predictions are valid *before* a substantial step is undertaken, then the threat of prepunishment has no deterrent effect on substantial steps.¹²

Overall, then, the move from a pure postpunishment legal system to a pure prepunishment legal system would make no difference to the willingly compliant and noncompliant, but it would create a significant hazard for the grudgingly compliant. To avoid prepunishment, the grudgingly compliant would need to change their preferences or engage in self-binding strategies that remove their future opportunities or abilities to act illegally. Self-binding strategies are easier to implement if prepunishment predictions are not valid until the offender takes a “substantial step.” If most of the grudgingly compliant are unable to become willingly compliant or to self-bind, then the move to a pure prepunishment system would result in their mass incarceration.¹³

2.2. Mixed Systems. Mixed systems combine both pre- and postpunishment threats. As with pure prepunishment systems, the effects of a mixed system are relevant only to the grudgingly compliant. The crucial variable is the difference in credibility between the prepunishment and postpunishment threats.

¹¹The substantial-step requirement, and requirements of a similar nature, feature in many modern statutes concerning attempt offenses. See the American *Model Penal Code* Section 5.01(2).

¹²Beebe (2008, 259–60) and Greene (ming) suggest that the “substantial-step” and related requirements for attempt and conspiracy charges have historically been at least partly motivated by a desire to prepunish — i.e., these requirements are at least sometimes seen as standards of evidence for claims about what a criminal would have done had they not been apprehended. If the point being made here is correct, then such requirements can be further motivated by the fact that they enable the threat of prepunishment to have some deterrent effect.

¹³Recall that we have been assuming a pure *preventative* prepunishment legal system. As an anonymous referee points out, pure non-preventative prepunishment systems would suffer from even more pronounced consequentialist problems. Suppose the grudgingly compliant are not adept at self-binding. In that case, the move from a postpunishment system to a non-preventative prepunishment system — say, where the prepunishment is some kind of corporal punishment — would result in an increase in the criminal disposition of the grudgingly compliant *and* it would fail to prevent their predicted crimes. From a consequentialist standpoint, such a system would be a disaster. (Another, but less severe, consequentialist worry for non-preventative prepunishment has been raised by Sorensen (2006, 173–4) and Kearns (2008, 252)).

Pre vs. Postpunishment %	Overall Punishment %	% Offender Risks
50, 50	75	50
80, 10	82	10
10, 80	82	80

FIGURE 2. Chance Profiles of Mixed Systems. The first column is the reliability of pre- and postpunishment, the second is the overall chance of punishment, and the third is the chance of punishment the offender is willing to risk at the time of decision.

The hazard for the grudgingly compliant is relatively low when the pre- and postpunishment mechanisms are each about 50% reliable. Imagine a modified MegaTech case in which half of the people who would steal are detected beforehand and prepunished, and half of the people who do steal (i.e., the ones who are not detected by the prepunishment mechanism) are detected afterwards and postpunished. In this case, there is a 50% chance of preventative prepunishment for those who would commit a crime, and for those missed by the prepunishment mechanism, there is a 50% chance of postpunishment. This mixed system results in the punishment of 75% of those who are willing to risk a 50% chance of postpunishment by stealing on Friday. The case is represented by the first row of Figure 2.

Now consider a case more heavily weighted toward prepunishment: a system of prepunishment that is 80% reliable combined with a system of postpunishment that is only 10% reliable. Thus, on Friday, there is only a 10% chance of postpunishment should you steal money, but there is an 80% chance of prepunishment for anyone willing to do this. In this situation, the mixed system results in the punishment of 82% of those who are willing to risk a 10% chance of postpunishment by stealing on Friday. This creates a hazard for the grudgingly compliant.

When the reliability of the postpunishment system is significantly greater than the reliability of the prepunishment system, the hazard for the grudgingly compliant virtually disappears. For example, if there is only a 10% chance of prepunishment but an 80% chance of postpunishment, a person who would steal money is willing to risk an 80% chance of postpunishment and faces only a 10% chance of prepunishment. In this situation, the mixed system results in the punishment of 82% of people who are willing to risk an 80% chance of postpunishment by choosing to steal on Friday. Thus, the prepunishment mechanism in this instance makes little difference to fate of the grudgingly compliant, since they are, by assumption, deterred by credible threats of postpunishment.

The model presented here thus suggests that as our abilities to prepunish increase, the hazard for the grudgingly compliant will become more apparent. Most fundamentally, the hazard concerns the relationship between the *overall chance of punishment* and the *chance of punishment a person risks by committing the criminal act*. The more credible the threat of prepunishment, and the less credible the threat of postpunishment, the more potential there is for these to diverge.¹⁴

¹⁴The models presented in this section aim to describe the deterrent effects (and lack thereof) of different kinds of prepunishment legal systems on members of a society. They do not aim to capture *all* the effects of prepunishment legal systems. For example, if the positive effects of preventing harm are strong enough, then they might overcome the negative effects stemming from a lack of deterrence (i.e., the increase in punishment of the grudgingly compliant). To measure all these effects against each other, a full consequentialist analysis would need to specify several additional variables, including the prevalence of grudging compliance and the

3. CONCLUSION

I have argued that prepunishment is deeply problematic for consequentialist reasons. Most importantly, prepunishment threats do not deter crimes at the time of decision but instead, at best, pressure would-be criminals to change their preferences over outcomes ahead of time or to self-bind. This reveals a previously unnoticed difference in the deterrent effects of pre- and postpunishment threats: postpunishment threats influence outcome-probabilities while prepunishment threats influence outcome-preferences and self-binding behavior. This calls for a more fine-grained approach to the study of deterrence and its use in consequentialist justifications of punishment.

If people are not adept at changing their preferences or self-binding, then we should expect that a wholesale move from post- to prepunishment in our legal systems would amplify criminal dispositions. A simple increase in the credibility of prepunishment threats, even within a postpunishment-oriented system, would also encourage criminal dispositions, by divorcing the overall chance of punishment from the chance of punishment conditional on the criminal act. Nevertheless, certain aspects of prepunishment remain desirable for consequentialists; most importantly, its use in preventing harm. Thus, two potential solutions to prepunishment's deterrence problem were discussed. The first, and simplest, would be to ensure that the credibility of postpunishment threats continues to remain vastly superior to the credibility of prepunishment threats. The second is to implement a strict "substantial-step" requirement for the validity of prepunishment convictions.

REFERENCES

- Beebe, H. (2008). Smilansky's alleged refutation of compatibilism. *Analysis*, 68(3):258–60.
- Gauthier, D. (1986). *Morals by Agreement*. Oxford University Press.
- Gibbard, A. and Harper, W. L. (1981). Counterfactuals and two kinds of expected utility. In Harper, W. L., Stalnaker, R., and Pearce, G., editors, *Ifs: Conditionals, Belief, Decision, Chance, and Time*, pages 153–90. D. Reidel.
- Greene, P. (2018). Success-first decision theories. In Ahmed, A., editor, *Newcomb's Problem*. Cambridge University Press.
- Greene, P. (Forthcoming). The real-life issue of prepunishment. *Social Theory and Practice*.
- Institute, A. L. (1981). *Model Penal Code*. Foundation Press (2002).
- Kavka, G. S. (1983). The toxin puzzle. *Analysis*, 43(1):33–6.
- Kearns, S. (2008). Compatibilism can resist prepunishment: A reply to smilansky. *Analysis*, 68(3):250–3.
- Levinstein, B. A. and Soares, N. (2020). Cheating death in damascus. *The Journal of Philosophy*, 117(5):237–66.
- McClellan, E. F. (1990). *Rationality and Dynamic Choice*. Cambridge University Press.
- Meacham, C. J. G. (2010). Binding and its consequences. *Philosophical Studies*, 149(1):49–71.
- Milevski, V. (2014). The utilitarian justification of prepunishment. *Polish Journal of Philosophy*, 8(1):25–5.
- New, C. (1992). Time and punishment. *Analysis*, 52(1):35–40.
- Parfit, D. (1984). *Reasons and Persons*. Oxford University Press.

feasibility of self-binding strategies. Such an analysis is of clear relevance to future research but beyond the scope of the current paper.

- Petersen, T. S. (2014). (Neuro)predictions, dangerousness, and retributivism. *The Journal of Ethics*, 18(2):137–51.
- Smilansky, S. (1994). The time to punish. *Analysis*, 54(1):50–3.
- Smilansky, S. (2007). Determinism and prepunishment: The radical nature of compatibilism. *Analysis*, 67(4):347–9.
- Soares, N. and Yudkowsky, E. (2018). Functional decision theory: A new theory of instrumental rationality. *arXiv: 1710.05060*.
- Sorensen, R. (2006). Future law: Prepunishment and the causal theory of verdicts. *Nous*, 40(1):166–83.
- Statman, D. (1997). The time to punish and the problem of moral luck. *Journal of Applied Philosophy*, 14(2):129–35.
- Wringe, B. (2012). Pre-punishment, communicative theories of punishment, and compatibilism. *Pacific Philosophical Quarterly*, 93:125–36.
- Zedner, L. (2007). Preventive justice or pre-punishment? The case of control orders. *Current Legal Problems*, 60:174–203.