Not Just Punishment:
The Overlooked Harms of the Criminal Legal System

by

Avital Fried

Submitted to the Philosophy Faculty

for the degree of

DPhil in Philosophy

Trinity 2023

University of Oxford

Word Count: 67,625

Abstract

This thesis contains six essays that explore the ways in which people who are accused of or convicted of crimes are *not just punished*. Rather, people are mistreated and systematically harmed far before a trial conviction and far beyond a standard prison sentence. Each essay critically evaluates an overlooked problem in the practices of criminal legal systems. Chapter 1 presents a novel challenge for the justification of **arrests**. Chapter 2 argues that the practice of crediting time served demonstrates that **pre-trial detention** is actually pre-trial punishment. Chapter 3 brings together the literatures on testimonial injustice and criminal **confessions** to investigate the epistemic harms that occur at and around trial. Chapter 4 demonstrates that defense-introduced **statistical evidence** is not subject to the standard reasons provided to exclude statistical evidence from trial. Chapter 5 argues that unique features of the **death penalty**, unlike standard incarceration, render capital punishment inherently disproportional. Finally, Chapter 6 demonstrates how uncontroversial premises lead us to the surprising conclusion that there is a stronger argument in favor of **compensation** for guilty people who are over-punished than for innocent people who are incorrectly punished.

Acknowledgements

First and foremost, thank you to my supervisors, Alex Kaiserman and Tom Sinclair, whose advice, guidance, and feedback have pushed me to think more deeply and more clearly. There are a number of people I would also like to thank for their comments on parts of this thesis, including Lily Moore-Eissenberg, Kevin Zhang, Anna Golova, Sarah Hirschfield, Livia Bokor, Tom Douglas, Ruth Chang, two anonymous reviewers at *Synthese*, and audiences at the 8th and 9th National Oxford Uehiro Ethics Competitions and at the 2022 Annual Conference of the Society of Applied Philosophy. I am appreciative of the financial support I received while writing this thesis from the British Marshall Scholarship and the Oxford Philosophy Faculty. I am deeply grateful to Noah for the time he spent reading the entire thesis, the incisive and helpful comments, and the incredible support throughout this whole thesis process and beyond. To my classmates at East Jersey State Prison and my students at Cook County Department of Corrections and Garden State Youth Correctional Facility: thank you for generously sharing your time, experiences, and opinions with me. You have taught me so much about the injustices of our criminal legal system, and you have shown me the humanity that survives within it. I dedicate this thesis to all of you.

Table of Contents

Introduction 1

The Challenge of Justifying Arrests 6

Pre-Trial Punishment: Demystifying Credit for Time Served 29

Criminal Confessions and Content-Sensitive Testimonial Injustice 66

Statistical Evidence and the Criminal Verdict Asymmetry 92

The Temporal Disproportionality of the Death Penalty 109

Compensating the Guilty over the Innocent 140

Conclusion 171

Introduction

Deep within a maximum-security prison in northern New Jersey, my class gathered to discuss the theory and practice of punishment. That semester, I was participating in an "Inside-Out" course that brought together students from my college and incarcerated students to learn with and from one another. One day, one of my incarcerated classmates, Brian, mentioned something that really struck me. He said that he views the prison population as the "muted minority." His voice, he explained, does not matter to society.

The incarcerated muted minority is silenced in many different ways. They are placed physically out-of-sight, locked behind layers of barbed wire in windowless buildings. The distance is amplified by restrictions on information entering and leaving prisons and jails. Everything is censored—internet is restricted, phone calls limited, letters monitored, and books screened. Brian explained that, for him, the muting is the denial of his political voice; he had been disenfranchised his entire life, as his voting rights had been stripped before he turned 18.

In the United States, there are approximately 2.3 million people in the muted minority.[1] Nearly one out of every 100 people in the US is incarcerated.[2] The number of people living behind bars is equivalent to the number of people living in Houston, Texas, the fourth largest city in the country.[3] Thousands of people are socially marginalized *before* their incarceration through systemic racism, *during* their incarceration through the silencing and monitoring practices in prisons and jails, and *after* their incarceration through disenfranchisement, post-conviction restrictions, and discrimination.

This thesis presents a collection of essays, each of which pertains to the ways in which people accused of and convicted of crimes are *not just punished*. From the moment of arrest to far beyond the end of a prison sentence, people who are accused of crimes are systematically harmed and mistreated in ways that extend beyond standard post-conviction incarceration.

Each essay in this thesis brings an overlooked harm to bear on a different philosophical puzzle or problem in the literature. Given that the philosophical literature has developed largely without these experiences and harms in mind, there are a variety of ways in which a fuller understanding of the criminal legal system can help clarify, refine, and challenge the philosophical scaffolding built around the practice and justification of punishment. In turn, through this philosophical lens, we can better understand the permissibility (or lack thereof) of these different practices and harms.

---

[1] Wagner and Bertram, ""What Percent of the U.S. Is Incarcerated?"
[2] Wagner and Bertram.
[3] "Houston, Texas Population 2023."

The essays in this thesis are organized around three primary themes: pre-trial harms, improper convictions, and off-limits punishments.

I. Pre-Trial Harms

The first two chapters of this thesis address two significant harms that occur before a person even reaches trial. In recent years, pre-trial harms have garnered attention in both political and scholarly discourse. In summer 2020, thousands of people took to the streets to protest police brutality against people of color.[4] In many jurisdictions in the US, activists pushed for the abolition of cash bail, trying to ensure that people are not detained pre-trial only because they are poor.[5] Though the issues of police brutality and cash bail are now part of mainstream political discussions, the focus on these issues obscures more foundational questions about pre-trial harms.

In Chapter 1, I identify a gap in the literature regarding the justification of the practice of arrest. Scholars have written extensively on common state practices that restrict personal freedom, such as pre-trial detention, punishment, quarantine, and preventive detention.[6] Yet, the practice of arrest, which is also liberty deprivation, has received almost no philosophical attention, despite its centrality and ubiquity. After outlining the harms of arrest, I argue that a justification of the practice must explain both why subpoenas are an insufficient alternative and why the harms associated with the practice of arrest are permissible. I then draw on the literature regarding the justifications of pre-trial detention to explore different possible justifications for the practice of arrest. I find that these justifications fail to justify using arrests as the default practice in criminal cases.

In Chapter 2, I turn to the practice of pre-trial detention. I first establish the two puzzles regarding crediting time served: the proportionality puzzle and the acquitted defendant puzzle. The proportionality puzzle argues that crediting time served leads to the systematic under-punishment of people who are detained pre-trial. The acquitted defendant puzzle points out that the pre-trial detention of convicted defendants appears to be treated differently than that of acquitted defendants, since convicted defendants receive credit while acquitted defendants do not receive anything for the same time spent in pre-trial detention. I argue that the practice of crediting time served provides a presumptive argument in favor of understanding pretrial detention as punishment, and I show that

[4] Buchanan, Bui, and Patel, "Black Lives Matter May Be the Largest Movement in U.S. History."
[5] See, e.g., Foderaro, "New Jersey Alters Its Bail System and Upends Legal Landscape."
[6] The past decade has seen renewed philosophical attention to pre-trial detention. For some pieces published in the early 2010s that reignited the discussion on pre-trial detention, see Duff, "Pre-Trial Detention and the Presumption of Innocence"; Ferzan, "Preventive Justice and the Presumption of Innocence"; Baradaran, "Restoring the Presumption of Innocence."

this understanding resolves both puzzles. I then demonstrate that pretrial detention fits standard definitions of punishment. Since pre-trial detention is punishment before a person has even been tried, the practice is straightforwardly impermissible. I end with a novel compensation and credit model for pre-trial punishment.

II. Improper Convictions

The third and fourth chapters of this thesis point out epistemic issues at trial. Fair trials are central to the legitimacy of state punishment, since proper convictions are necessary for permissible punishment. Ideally, jurors are good epistemic agents who properly weigh the evidence to reach the right conclusion. But the typical trial is far from that ideal practice—jurors may improperly weigh testimony, judges may exclude important evidence, and attorneys may misrepresent the facts. With the legitimacy of a state's system of punishment on the line, creating the conditions for proper convictions is of the utmost importance.

In Chapter 3, I bring together the literature on epistemic injustice and the literature on criminal confessions to make two interconnected contributions. First, I draw on a case of excessive reliance on false confessions to refine the concept of testimonial injustice. I argue for an understanding of testimonial injustice that clarifies (1) the types of prejudices involved, (2) the possibility for credibility excess testimonial injustice, and (3) what I term the content-sensitive nature of testimonial injustice. Second, with this clarified understanding of testimonial injustice, I demonstrate that criminal defendants—both those who are factually innocent and those who are factually guilty—are vulnerable to testimonial injustice at trial because of prejudicial beliefs held against criminal defendants. I demonstrate that instances of testimonial injustice at trial increase the chances of wrongful conviction, undermine the fairness of trials, and indirectly influence plea bargaining.

In Chapter 4, I begin with the literature on the proof paradox, which pertains to the exclusion of statistical evidence in criminal trials. I point out an important and previous unnoticed asymmetry between the epistemic norms of convictions and those of acquittals. Given this asymmetry, I argue that statistical evidence can be sufficient for an acquittal, even though statistical evidence cannot be sufficient for a conviction. I demonstrate how the exclusion of defense-introduced statistical evidence can lead to convictions without sufficient evidence and even to convictions of innocent defendants.

III. Off-Limits Punishments

The last two chapters of this thesis consider the over-punishment of people who have committed crimes. Philosophers and legal scholars have written extensively on what, if anything, justifies a system of punishment. Typically, theories of punishment point to a difference between guilty and innocent people in order to answer this question. We tend to think that it would be impermissible for the state to round up a random group of citizens and incarcerate them for months or years. Yet, many people think it is permissible to round up a group of citizens who have committed crimes and incarcerate them for months or years. The focus on differentiating between the guilty and the innocent, however, has left the issues regarding over-punishing the guilty underexplored.

Chapter 5 presents a novel anti-death penalty argument. Since much of the pro-death-penalty literature relies on retributivist concepts, many of the anti-death penalty arguments attack basic commitments of a pro-death-penalty retributivist, including the notion that a person can deserve death. Even granting the pro-death-penalty retributivist all of her commitments, I demonstrate that the death penalty, unlike other forms of standard punishment, is inherently disproportional. First, I contend that there is a necessary tradeoff between the need for a thorough appellate process that ensures desert and the disproportional harm experienced by those incarcerated on death row. Second, I demonstrate how the harm of execution makes it a necessarily disproportional penalty.

In Chapter 6, I consider the question of compensation for over-punishment. I begin with the commonly-held view that the state should compensate innocent people who are incorrectly punished. To address the question of compensation for guilty people who are over-punished, I turn to the literature on the exculpatory power of false beliefs. First, I demonstrate that there is a stronger argument for compensation when the state's action is based on a false moral belief than when the state's action is based on a false factual belief. Second, I show that incorrectly punishing an innocent person is typically a wrong action based on a false factual belief, while punishing a person beyond the limits of permissible punishment is typically a wrong action based on a false moral belief. Finally, I arrive at the surprising conclusion that compensation for over-punishing a guilty person is actually more defensible than compensation for incorrectly punishing an innocent person.

This thesis aims to demonstrate that bringing together the overlooked harms of the criminal legal system and the relevant philosophical literature is necessary, productive, and mutually informative.

Bibliography

Baradaran, Shima. "Restoring the Presumption of Innocence." *Ohio State Law Journal* 72, no. 4 (2011): 723–76.

Buchanan, Larry, Quoctrung Bui, and Jugal K. Patel. "Black Lives Matter May Be the Largest Movement in U.S. History." *The New York Times*, July 3, 2020, sec. U.S. https://www.nytimes.com/interactive/2020/07/03/us/george-floyd-protests-crowd-size.html.

Duff, R. A. "Pre-Trial Detention and the Presumption of Innocence." In *Prevention and the Limits of the Criminal Law*, edited by Andrew Ashworth, Lucia Zedner, and Patrick Tomlin, 115–32. Oxford University Press, 2013. https://doi.org/10.1093/acprof:oso/9780199656769.003.0007.

Ferzan, Kimberly Kessler. "Preventive Justice and the Presumption of Innocence." *Criminal Law and Philosophy* 8, no. 2 (June 1, 2014): 505–25. https://doi.org/10.1007/s11572-013-9275-0.

Foderaro, Lisa W. "New Jersey Alters Its Bail System and Upends Legal Landscape." *The New York Times*, February 7, 2017, sec. New York. https://www.nytimes.com/2017/02/06/nyregion/new-jersey-bail-system.html.

"Houston, Texas Population 2023." Accessed June 11, 2023. https://worldpopulationreview.com/us-cities/houston-tx-population.

Wagner, Peter, and Wanda Bertram. "'What Percent of the U.S. Is Incarcerated?' (And Other Ways to Measure Mass Incarceration)." Prison Policy Initiative, January 16, 2020. https://www.prisonpolicy.org/blog/2020/01/16/percent-incarcerated/.

The Challenge of Justifying Arrests

Every three seconds, a person is arrested in the United States. Arrest is a widespread practice in many, if not all, jurisdictions. It often marks the beginning of a sustained period of incarceration, which includes pre-trial detention and post-trial punishment. This paper establishes a novel challenge for justifying the practice of arrest. I first explain the serious harms of arrest and then contend that subpoenas appear to be a less harmful alternative. I find that the state must either justify its use of a harmful practice or abandon the practice in favor of the less harmful alternative. Drawing on the pre-trial detention literature, I extend the challenge by demonstrating that the most obvious contenders for potential justifications for arrest only cover a much narrower subset of cases than those that we find in practice. Thus, a serious challenge remains for justifying the default practice of arrest in criminal cases.

Introduction

On May 25, 2020, George Floyd was murdered during an arrest for allegedly using a counterfeit $20 bill. A Minneapolis police officer knelt on Floyd's neck for almost nine minutes. Protests erupted around the world, criticizing the pattern of law enforcement's excessive use of force—often deadly force—against people of color.[7]

The murder of Floyd demonstrated that there is something clearly wrong with the current state of policing in the US. In response, public attention and advocacy focused on the fatal nature of arrests. Due to the numerous stories similar to Floyd's, calls for body cameras, de-escalation techniques, and defunding the police increased.[8] These calls, however, obscured a more basic question: Even if all arrests were non-violent, is the practice of arrest justified in the first place?

Approximately 12 million people are arrested in the US every year, meaning that a person is arrested every three seconds on average.[9] Across the world, arrests often mark the beginning of a sustained period of incarceration, which includes pre-trial detention and post-trial punishment. From the time a person is forcibly taken into custody until the person is charged with a crime, the person is

[7] Hill et al., "How George Floyd Was Killed in Police Custody." All future references to this case are from here.
[8] Alter, "Activists Want to End Police Violence. But Disagree on How to Do It."
[9] Neusteter and O'Toole, "Every Three Seconds: Unlocking Police Data on Arrests."

considered under arrest. Indeed, this initial period can influence, and may even dictate, whether the person is ultimately found guilty and the length of the sentence.[10]

Despite the centrality and ubiquity of arrests, there is a surprising lack of attention among both activists and scholars to this issue. Activists have focused on reducing pre-trial detention by pushing for the abolition of cash bail, a system that detains people pre-trial due to their inability to provide a financial guarantee that they will attend trial.[11] Yet, the 24–72 hours that precede the official start of pre-trial detention have been largely overlooked. Practices that are similarly widespread and restrictive of personal freedom—such as pre-trial detention, punishment, quarantine, and preventive detention—have all received significant philosophical attention.[12] Some legal scholars have criticized the current scale of arrests in the US, and others have noted that arrests may be not be permissible in certain cases.[13] Even with these discussions of arrests, however, there is a lack of attention to the moral stakes of the practice more broadly.[14]

This paper presents a robust challenge for justifying the widespread practice of arrest. In §1, I briefly define the practice. In §2, I present the **Subpoena Alternative Argument**, which contends that the burden of justifying arrests falls on the state. I explain that the state must justify both why subpoenas are an insufficient alternative and why the harms of arrest are permissible in the first place. In §§3–4, I turn to the philosophical literature on pre-trial detention for potential justifications of the practice of arrest. I demonstrate that the standard justifications for depriving people of their liberty pre-trial can only cover a much narrower subset of instances of arrest than those that we find in practice.

---

[10] For studies that show that detention in police custody, even for only a few days, can increase likelihood of conviction and sentence length, see, e.g., Heaton, Stevenson, and Mayson, "The Downstream Consequences of Misdemeanor Pretrial Detention"; Digard and Swavola, "Justice Denied: The Harmful and Lasting Effects of Pretrial Detention"; Dobbie, Goldin, and Yang, "The Effects of Pretrial Detention on Conviction, Future Crime, and Employment"; Stevenson, "Distortion of Justice."

[11] See, e.g., "The State of Bail: A Breakthrough Year for Bail Reform."

[12] Given the robust philosophical attention to these issues, only a sampling of pieces on all four topics can be provided here. For pre-trial detention, see Duff, "Pre-Trial Detention and the Presumption of Innocence." For punishment, see Altman, *A Theory of Legal Punishment*. For quarantine, see Douglas, "Is Preventive Detention Morally Worse than Quarantine?" For preventive detention, see Husak, "Preventive Detention as Punishment?: Some Possible Obstacles."

[13] See, e.g., Harmon, "Why Arrest?"; Engel et al., "Alternatives to Arrest"; Gershowitz, "Justice on the Line."

[14] Harmon recognizes this problem as well: "Generous law enforcement authority to arrest exists, largely unexamined…we take the power to arrest for granted." See Harmon, "Why Arrest?," 309.

§1. Defining Arrest

Let us begin by articulating what precisely is meant by arrest. Though the practice of arrest is widespread, the precise details of the practice vary between jurisdictions. For the sake of clarity, I take the practice of arrest in the US as the focus of this paper.[15]

The term "arrest" is not defined in codified US federal law. Instead, US Supreme Court opinions over time have developed different explanations as to what constitutes an arrest.[16] In Miranda v. Arizona (1966), the court describes a situation of arrest as one in which "a person has been taken into custody or otherwise deprived of his freedom of action in any significant way."[17] Since this description is somewhat ambiguous, let us consider various instances of arrest in order to build a working definition for this paper.

Typically, when a person is arrested, she is forcibly taken into police custody and detained until she is officially charged with a crime. At that point, a judge decides whether the person may be offered bail. If the person is denied bail or if the person cannot pay the required amount to post bail, then she may be detained pre-trial. The stretch of time between when a person is first forcibly taken into police custody until she is charged with a crime is typically 24–72 hours, though there are instances in which a person may be detained without being charged for far longer.[18] There are also instances in which a person is taken almost directly to court, such that she may be in police custody for only a few hours.

While some arrests are carried out with warrants, other cases of arrest are warrantless. In a warrant arrest case, a person may have already been charged with a crime. The prosecutor secures a warrant prior to arrest by demonstrating to a judge or a grand jury that there is probable cause to believe that the person committed a crime. The arrest warrant then states the legal justification for the arrest, namely the crimes that the arrestee has allegedly committed. Arrest warrants are most typically used in high profile cases or following a long-term investigation. If, for instance, the police have concluded that A likely committed a murder after a thorough investigation, then a prosecutor would submit the evidence to a judge and receive a warrant for A's arrest. The police would then locate A—regardless of whether she is at work, home, or elsewhere—and arrest her.

---

[15] The arguments presented here would similarly apply to other jurisdictions with the same or similar arrest practices and stated justifications.

[16] For an overview of the jurisprudence on the definition of arrest, see Clancy, "What Constitutes an Arrest within the Meaning of the Fourth Amendment."

[17] Miranda v. Arizona, 384 U.S. 436 (1966) at 444.

[18] For instance, the US detained people in Guantanamo Bay for years before, and sometimes without ever, charging them with a crime. See Rosenberg, "Pentagon's Repatriation of Algerian Leaves 30 Prisoners at Guantánamo."

A warrantless arrest—the more common form—occurs when a police officer believes that a person has committed a crime or is committing a crime but does not first go to a judge or grand jury for a warrant. In Floyd's case, for instance, the police were called to the location because the store clerk thought he had received a counterfeit bill. Though the officers did not have a warrant, they arrested Floyd on the basis that they had probable cause (i.e., a reasonable belief) that Floyd had committed the crime of using a counterfeit bill.

Though arrests may take many forms, the common factor is that once the person is brought before a judge, she is no longer under arrest but rather is considered a criminal defendant. For the purposes of this paper, our working definition is as follows:

> **Arrest**: An arrest is a deprivation of liberty that is imposed by a legal authority in response to the alleged commission of a crime and that precedes a person's official introduction into the criminal legal system.

There are two further types of liberty deprivations, which, although resembling arrests, do not fall into our definition of arrest. First, there are instances in which people are taken into police custody in which the deprivation of liberty is not in response to an alleged crime.[19] For instance, if a person is having a mental health crisis and thereby poses a threat to herself or others, police may forcibly take her into custody to bring her to a hospital. Though such cases similarly involve a police-imposed liberty deprivation, they are not triggered by alleged crimes, nor do they introduce the person into the criminal legal system. Thus, these liberty deprivations are importantly distinct from standard cases of arrest.

Second, a citizen's arrest does not fall into our definition of arrest, since it is not imposed by the police or another legal authority. Some jurisdictions empower non-police officers to arrest others if they have reason to believe that a person is committing a crime and would flee the scene before the police arrive. This practice is highly problematic for a number of reasons, some demonstrated by the killing of Ahmaud Arbery during a citizen's arrest in Georgia in February 2020.[20] Though such cases similarly involve the deprivation of liberty in response to a crime, the deprivation is not imposed by the police or another legal authority.

---

[19] Some of the legal literature on arrests focuses its attention on forcible police custody in these instances as examples of problematic "arrests." (See, e.g., Engel et al., "Alternatives to Arrest.") Understood in this way, the term "arrest" becomes overly broad, since the justification and motivation for taking a person into custody in the middle of a mental health crisis is drastically different than the colloquial and typical notion of arrest.

[20] See, e.g., Valentine, "What Is the Citizen's Arrest Law in the Trial over Ahmaud Arbery's Death?"; Flanders et al., "The Puzzling Persistence of Citizen's Arrest Laws and the Need to Revisit Them."

There is another important distinction between two kinds of arrests that both meet our working definition: some arrests are post-crime, while other arrests are mid-crime.[21] A post-crime arrest could occur with or without a warrant. If a person allegedly committed the crime many days prior to her arrest, then the arrest would likely be a warrant arrest. If the person allegedly committed the crime just a few minutes before her arrest, then the arrest would be warrantless. Regardless of the timing, the important feature in both of these situations is that the crime has already been committed. In contrast, a mid-crime arrest stops an ongoing event and prevents the completion of the crime's commission.

For the purposes of this paper, I am setting aside the category of mid-violent-crime arrests for a few reasons. Arrests carried out in the middle of violent crimes may have different aims and may be justified on different grounds than those of post-crime arrests. Though a person arrested mid-crime will likely enter the criminal legal system following the deprivation of liberty, the primary purpose of liberty deprivation is not to get the person before a judge. Rather, the purpose is to protect other people. If there is an active shooter, for instance, an arrest prevents the person from harming other people. Thus, a mid-violent-crime arrest may be justified on other-defense grounds. In that way, it may be more akin to the police forcibly taking a person experiencing a mental health crisis to a hospital.

Now that we have settled our definition of arrest, let us turn to the challenge of justifying the practice.

§2. Subpoena Alternative Argument

Arrest functions as the gateway into the criminal legal system. Yet, the US also uses a different way to bring people to court: subpoena. Only criminal defendants are arrested; all others, such as civil defendants or witnesses in criminal trials, are subpoenaed if they are required to appear in court. This section presents the **Subpoena Alternative Argument**, which proceeds as follows:

P1. Arrests harm people.

P2. Subpoenas are less harmful than arrests.

P3. Arrests and subpoenas fulfill the same function.

P4. If a state's practice A harms people, and there is an alternative practice B that fulfills the same function and is less harmful, then the state should use practice B instead of practice A.

C: The state should use subpoenas instead of arrests.

[21] In theory, an arrest could be pre-crime if a person is planning or attempting to commit a crime. But, planning to commit a crime is itself a crime, so the arrest would be mid-crime for an attempt or conspiracy crime.

It would appear that the best path forward for the proponent of arrests is to explain why subpoenas and arrests do not fulfill the same function (i.e., to deny P3). In this section, I defend the premises of the argument, beginning with the overlooked harms of arrest.

2.1 Harms of Arrest

In this section, I outline a variety of harms, both potential and inevitable, involved in the practice of arrest. Though the list in this section is not meant to be exhaustive, it demonstrates how the practice of arrest is seriously harmful (P1).

The harms most commonly associated with arrests are harms from police force and brutality. As discussed above, instances of deadly force during arrests have garnered attention worldwide. On July 17, 2014, Eric Garner was killed during an arrest when a police officer put him in a fatal chokehold. The deaths of George Floyd and Breonna Taylor during arrests in 2020 sparked a resurgence of the Black Lives Matter movement.[22] Though these are some of the most famous cases, a 2019 study found that there are 375–496 arrest-related deaths in the US per year.[23] Notably, this statistic only captures those who are killed; the number of people per year who experience non-fatal police violence during an arrest is likely significantly higher.[24]

For this paper, however, let us put aside the crisis of police brutality.[25] Though this inexcusable violence needs to be addressed, the focus on the violent aspects of arrests obscures the more basic question of whether the practice, even when carried out correctly, is justified.

Even when properly administered, an arrest is harmful. It is a coercive practice that deprives a person of her liberty. Given the seriousness of state-imposed deprivations of liberty, there is extensive literature on what, if anything, justifies state-enforced losses of liberty in a variety of cases. For instance, scholars have written on state-enforced quarantines, preventive detentions, and prison sentences.[26] In all of these areas, philosophers tend to assume there is a presumptive right to liberty and thus attempt to articulate what would justify the *denial* of liberty in these cases. Since the right to liberty is seen as a foundational right worth protecting, it is clear that a denial of liberty is a serious harm.

---

[22] Demby and Marisol Meraji, "A Decade of Watching Black People Die."
[23] Ordway, Whibey, and Walter Kille, "Deaths in Police Custody in the United States."
[24] Weichselbaum et al., "Violent Encounters with Police Send Thousands of People to the ER Every Year."
[25] Importantly, people also suffer from police violence and brutality while in custody, both while they are under arrest and while they are detained pre-trial. See, e.g., Katz, "Chokeholds, Chemical Spray, and Head Strikes."
[26] See, e.g., Duff, "Pre-Trial Detention and the Presumption of Innocence"; Douglas, "Is Preventive Detention Morally Worse than Quarantine?"; Ferzan, "Preventive Justice and the Presumption of Innocence."

There are also a variety of more subtle ways in which arrestees may be harmed. First, an arrested person may lose her ability to work and earn money. The loss of wages can affect the arrested person's family. The person may not be able to communicate with family or attend important life events. Recently, philosophers have taken to considering different ways in which prison sentences can indirectly harm a person's family.[27] The loss of liberty involved in arrest, too, can lead to other harms to the arrestee and her family.

Second, a person may suffer from reputational damage as a result of an arrest. Scholars have recognized that arrest is often conflated with guilt.[28] When a person is arrested, others often assume (whether correctly or not) that the person is guilty of committing a crime. In popular media, an arrest is often closely associated with a public demonstration of guilt.[29] Arrests can happen in public places and venues, including at a person's workplace. The public nature of an arrest can damage a person's reputation. Indeed, reputational harm can have long-lasting effects.[30] Numerous studies have shown that an arrest alone (without a charge or a conviction) often leads to future employment discrimination.[31] People may be required to disclose an arrest record when applying for jobs, educational programs, and housing. As such, a mere arrest can have long-term negative impacts on a person's life.

Third, an arrested person suffers from a loss of privacy. People who are arrested are required to provide biophysical data to the police. It is common practice to take mugshots and collect fingerprints of those arrested. Other personal information (e.g., height, weight, date of birth, eye color) may be entered into a police database. Typically, people must consent to sharing private information and data with the state. Following an arrest, however, no such consent is expected.

Fourth, the deprivation of liberty may hinder a person's ability to properly prepare for court. While a person is detained for 24–72 hours, it is more difficult to get in touch with an attorney, and the ability to prepare evidence and arguments for a bail hearing is limited.[32] Indeed, studies have shown

---

[27] Zachary Hoskin is perhaps the philosopher currently most focused on the issue. See, e.g., Hoskins, "Punishment's Burdens on the Innocent"; Hoskins, "Collateral Legal Consequences and Criminal Sentencing"; Hoskins, "Guilty Pleas, Sentence Reductions, and Non-Punishment of the Innocent." For other pieces on the issue, see, e.g., Condry and Minson, "Conceptualizing the Effects of Imprisonment on Families"; Lippke, "Punishment Drift."

[28] Roberts, "Arrest as Guilt."

[29] Hope, Memon, and McGeorge, "Understanding Pretrial Publicity."

[30] Geoffrey Scarre argues that reputational damage is so severe that it should mitigate a person's sentence. See Scarre, "Corrective Justice and Reputation."

[31] See, e.g., Calvani, "Discrimination on the Basis of Arrest Records"; Barnes, "Study Says 'Ban the Box' Policies Hurt (Not Help) Young Minority Male Job Seekers"; Bushway et al., "Barred from Employment"; Geffen, "The Collateral Consequences of Acquittal."

[32] The bail conditions, including the amount required for release, are typically determined by a judge. If a person has not had the chance to speak to a lawyer, she may be offered less favorable bail terms. Studies have shown that the inability to

that the time spent incarcerated while under arrest can impact both the likelihood of conviction and sentence length.[33] Moreover, in the context of forced confinement, arrestees are often questioned and interrogated about the alleged crime. Many jurisdictions recognize a person's vulnerability when subject to police suspicion while in police custody. In the US, a person possesses a right against self-incrimination, and there are a variety of safeguards put in place by courts to ensure that people know that they do not have to talk to the police (e.g., Miranda rights).[34] Despite the right to remain silent, the combination of forced police custody and suspicion can lead to different types of harms. For instance, Jennifer Lackey writes about the epistemic harms involved in extracting testimony in this context.[35] She argues that an agent suffers from *agential testimonial injustice* when her agency is bypassed in the creation of testimony due to coercion, stress, and other factors.

The practice of arrest, even in its proper form, is a coercive deprivation of liberty. It also inherently risks the infliction of a variety of other potential harms, including some that far outlast the period during which a person is under arrest.

2.2 Subpoenas

Subpoenas are far less harmful than arrests (P2). Since this premise is fairly obvious, I will only briefly run through the argument. First, when a person is subpoenaed, there is no physical restriction of liberty. The person is not put into handcuffs or physically restrained. The person can continue to work in the time between receiving the subpoena and appearing in court. Moreover, the person remains at home with her family and community. The only liberty restriction is not a physical one—the state requires that the person appear in a certain place at a certain time, a restriction not dissimilar from many liberty restrictions that one may otherwise experience.[36]

Second, subpoenas are less reputationally harmful than arrests. Subpoenas are issued in an envelope, which makes them more private. If a subpoena is sent by mail, it can be received in the privacy of one's home. Even if a subpoena is delivered in a public place, it still causes less reputational

---

post bail can negatively impact a person's final sentence. See, e.g., Dobbie, Goldin, and Yang, "The Effects of Pretrial Detention on Conviction, Future Crime, and Employment"; Heaton, Stevenson, and Mayson, "The Downstream Consequences of Misdemeanor Pretrial Detention"; Stevenson, "Distortion of Justice."

[33] For studies that show that detention in police custody, even for only a few days, can increase likelihood of conviction and sentence length, see, e.g., Digard and Swavola, "Justice Denied: The Harmful and Lasting Effects of Pretrial Detention"; Leslie and Pope, "The Unintended Impact of Pretrial Detention on Case Outcomes."

[34] Miranda v. Arizona, 384 U.S. 436 (1966).

[35] Lackey, *Criminal Testimonial Injustice*.

[36] The most similar example of such a liberty restriction is jury duty, since the state similarly requires people to appear in court at a designated time.

harm, as subpoenas are used for anyone requested in court, including witnesses and civil defendants. Subpoenas are also only made public if there is reason to do so. A past subpoena would therefore have no influence on a person's future job prospects, educational opportunities, or housing options.

Third, when a person is subpoenaed, she is not required to provide biophysical data to the state. Fourth, with the notice provided by a subpoena, a person is likely to be more prepared for court. The person can more easily consult with lawyers and put together a stronger case. Finally, there is far less risk of a variety of other harms, since the person is not in the state's custody for an extended period of time.

Arrests and subpoenas both fulfill the primary function of bringing a person to court (P3). In criminal cases, arrests are typically used to introduce a defendant into the criminal legal proceedings. Everyone else is brought to court via subpoenas, including defendants in civil cases. Subpoenas are also used to summon witnesses, acquire documents, and compel people to testify. Though there may be other functions fulfilled by arrests and subpoenas, it seems that arrests and subpoenas fulfill the same primary function.

Now let us turn to the final premise: if a state's practice A harms people, and there is an alternative practice B that fulfills the same function and is less harmful, then the state should use practice B instead of practice A (P4). Generally, the state is expected to take reasonable steps to avoid harming people. In reality, a state will inevitably inflict harm while implementing policy and pursuing different aims. Yet, the state is expected to try to mitigate that harm, even if it is just an unintended byproduct of another goal. If the state should minimize the harms it imposes, then, when presented with two options to achieve the same function, the state should opt for the less harmful option.

Since subpoenas are a less harmful alternative to arrests, it seems that the state should prefer subpoenas as the default practice for bringing people to court. For a proponent of arrests, it appears that the best rebuttal to this conclusion is to deny P3. P3 states that arrests and subpoenas fulfill the same primary function. If the proponent of arrest can identify another primary function that can only be fulfilled by arrests and not by subpoenas, then P3 would not be true, and the conclusion of the **Subpoena Alternative Argument** would not follow.

Even if the proponent of arrest can successfully demonstrate that another primary function can only be fulfilled by arrests, there would still be the further question of whether the additional function justifies the harms of arrest. For instance, if the use of arrests instead of subpoenas meant that the state could employ more people, then arrests would serve a function that subpoenas do not. The question, then, would be if creating more jobs justifies the significant harms of arrest.

Thus, we have arrived at a challenge for the proponent of arrests. In order to successfully vindicate the practice, the justification needs to explain why:

(1) arrests fulfill an additional primary function that subpoenas cannot, and

(2) the harms of arrest are permissible, given that additional primary function.

In the coming sections, I propose some potential justifications on behalf of a proponent of arrests and evaluate whether the justifications meet these criteria. Given the lack of literature on arrests, I turn to the legal and philosophical literature on pre-trial detention, since arrests similarly deprive people of their liberties before they have been tried and convicted.

§3. Ensuring Justice

In the US, a person may be detained pre-trial to ensure the "appearance of the person as required and the safety of any other person and the community."[37] Thus, there are two standard approaches to justifying pre-trial detention: ensuring a proper trial and preventing harm to the community. In the coming section, I explore how the first justification, ensuring a proper trial, may or may not justify arrests.

3.1 Obstructing Justice

A first possible explanation as to why arrests are justified is that, if not arrested, a person may "obstruct justice" by fleeing the jurisdiction, tampering with evidence, or intimidating witnesses. As such, the person must be taken into custody by surprise so that she does not have the chance to obstruct justice.

According to this explanation, the goal of an arrest is to ensure that the person arrives at trial and does not tamper with evidence. Of course, those concerns exist in civil cases as well, and yet subpoenas are used in those contexts. Why not use subpoenas to ensure people show up in court for criminal cases?

One possible distinction between civil and criminal cases is the penalty at stake. The argument is as follows: a person facing a criminal trial, and thus the possibility of incarceration, may want to do whatever is in her power to avoid incarceration. As such, the person may attempt to flee the jurisdiction in an effort to avoid standing trial or tamper with evidence to reduce the chances of conviction. Because a prison sentence is at stake, there is a higher likelihood that the person may obstruct justice. As such, the subpoena would serve to give the defendant a warning, providing her

[37] 18 U.S. Code §3142 Release or Detention of a Defendant Pending Trial.

with time to flee or tamper with evidence. Thus, an arrest is needed to take the defendant into custody as soon as she is notified of her impending trial.

In civil cases, the argument continues, the stakes are different. Because the penalty is not incarceration, the defendant may be less likely to flee or to tamper with evidence. As a result, a subpoena, which allows the defendant to remain free in between the notice and the hearing, is sufficient. Thus, in "high stakes" cases, arrests are necessary to ensure that people appear in court to be charged with a crime, while in "low stakes" cases subpoenas are sufficient, since people will likely appear in court regardless. According to this argument, arrests fulfill a primary function that subpoenas cannot (i.e., preventing the obstruction of justice), thus satisfying (1).

I have two responses to this argument. First, I contend that there is no relevant categorical difference between criminal and civil cases. Second, I explain how subpoenas could be effectively used in criminal cases in the same way that they are used in civil cases, thus refuting the argument that arrests fulfill a primary function that subpoenas cannot.

First, the stakes are not categorically higher in criminal cases. There are certainly many high-stakes criminal cases, but there are also many high-stakes civil cases. For instance, there are civil cases in which a person has her home and life savings on the line. Alternatively, there are instances in which a person's reputation could be harmed by a finding in a civil case. Currently, in the US, there are a number of high-stakes civil cases pending regarding former President Trump's companies and employees. In these cases, millions of dollars, the solvency of the companies, and Trump's reputation are all on the line. Yet, subpoenas are used in these cases. Similarly, subpoenas have been used to call witnesses to testify in front of the US Congress, including to answer questions about the January 6th insurrection at the US Capitol. Thus, there are many examples of high-stakes, non-criminal situations in which subpoenas are used. Conversely, there are also low-stakes criminal cases; a person could be facing a very short sentence, just probation, or even a community service mandate.

Moreover, there is no categorical difference between the potential penalties involved in criminal and civil cases. Incarceration may be issued as a penalty for people failing to pay fines, and people may be monetarily penalized in criminal cases. For instance, a woman in the US was arrested and jailed 19 times, totaling 67 days of incarceration, for failing to pay a parking ticket, which is a civil offense.[38] Similarly, people who are convicted of criminal offenses are often required to pay fines in addition to their incarceration.[39] Since many criminal defendants are poor, they pay these fines by

[38] "A Debtor's Prison."
[39] "Paying for Your Time."

working while incarcerated.[40] They never receive those wages, and instead the money goes towards their fines. (This practice has been criticized for effectively amounting to slavery.[41]) Thus, the line that separates incarceration and fines into criminal and civil arenas, respectively, does not actually exist. As a result, justifying the use of arrests instead of subpoenas by appealing to the higher stakes in criminal cases does not stand up to scrutiny.

Second, there is an existing system to ensure compliance when subpoenas are issued that could also be used to ensure criminal defendants' compliance with subpoenas. If a person does not comply with a subpoena in a civil case, the person may be arrested and then face additional penalties. Similarly, in a criminal case, a person could be arrested and then face additional penalties if she does not comply with a subpoena. These penalties—in both civil and criminal cases—would serve to deter people from ignoring subpoenas.

If the state can demonstrate that a person is likely to fail to comply with a subpoena, the evidence can be presented to a judge who could then issue a warrant allowing the state to preemptively seize property in order to secure the potential penalty. In the context of cases with monetary penalties, suppose that the state has good reason to believe that a person will attempt to move her assets abroad, out of reach of the US government. If so, the state may seize a person's assets through a freezing injunction (previously known as a mareva injunction). After the judgment, the assets are either returned to the person or given to the winning party. The same process could exist for cases whose penalty is incarceration. If the state has good reason to believe that a person will attempt to flee the jurisdiction, then the state could present that evidence to a judge in order to obtain a warrant to arrest the person. Arresting the person would be meant to ensure that the state will be able to impose the penalty, namely incarceration, if the person is found guilty.[42]

In this scenario, the arrest would be serving a function that a subpoena cannot serve. When used for subpoena non-compliance, arrest would be ensuring that the state will be able to impose a penalty should the person be found guilty. The scope of this justification, however, is fairly narrow. In cases with monetary penalties, assets are not preemptively seized. Rather, the default practice is to issue a subpoena and let the case run its course. The exception is when there is evidence that a person has the capacity to move her assets beyond the reach of the US government. The evidence must then

---

[40] Menendez et al., "The Steep Costs of Criminal Justice Fees and Fines."
[41] See, e.g., Sainato, "'Slavery by Any Name Is Wrong.'"
[42] The parallels continue also when a subpoena or an arrest fails; if there is a civil judgment and monetary penalty but the assets are unreachable, then an injunction can be issued that would allow the state to seize the assets if they later become within the state's reach. Similarly, if a person flees the jurisdiction, the person can be tried *in absentia*. If the person later becomes within the state's reach, the person may be arrested and incarcerated, without a further trial.

be presented to a judge in order for a warrant to be issued. Similarly, if subpoenas were used for criminal defendants, the default would be to simply issue a subpoena. Arrest would be for exceptional cases in which the state could demonstrate to a judge that the person was likely to flee or had already failed to comply with a subpoena.

One could argue that all criminal defendants are likely to flee. In the US, one of the justifications of pre-trial detention is that the detention ensures a person's presence at trial.[43] Given the cash bail system, it seems that courts may be concerned that many, if not all, criminal defendants will flee. Since a court wants to ensure that defendants appear at trial, the court demands that the defendant pay a collateral. If the defendant appears at trial, then the court returns her money. If the defendant does not appear at trial, she forfeits the money she originally gave the court. The fact that a defendant must monetarily guarantee her attendance beforehand seems to indicate that the court assumes that many criminal defendants would otherwise flee. If this were so, then perhaps arrests are justified because they appear to fulfill another primary function that subpoenas cannot (i.e., preventing criminal defendants, all of whom the courts suspect of fleeing, from doing so).

The movement to abolish cash bail, and its adoption in a number of jurisdictions, indicates that courts are not concerned that all criminal defendants will flee.[44] Jurisdictions that have abolished cash bail have not decided to detain *all* criminal defendants.[45] Rather, they now *de facto* release defendants to await trial at home unless the defendant qualifies for one of the justifications for pre-trial detention (e.g., there is strong evidence that the person is likely to flee the jurisdiction).[46] Thus, the ways in which jurisdictions have responded to calls to abolish cash bail demonstrate that there is not a presumption that all criminal defendants are likely to flee.

One implication of this view is that the harms of pre-trial detention and arrest would primarily fall on the rich, not on the poor. In its current form, pre-trial detention largely burdens the poor, since they are financially unable to post bail. If people are only arrested and detained pre-trial when they pose a flight risk, then most of the people arrested and detained would be those with the means to escape the jurisdiction of the US. Fleeing the country likely requires significant funds, a non-public means of transportation, and a destination that does not have an extradition treaty with the US. Thus, the people who are most likely to flee are the rich.

---

[43] 18 U.S. Code §3142 Release or Detention of a Defendant Pending trial.
[44] See, e.g., "The State of Bail: A Breakthrough Year for Bail Reform."
[45] See, e.g., Foderaro, "New Jersey Alters Its Bail System and Upends Legal Landscape."
[46] See, e.g., Merkl, "New York's Upcoming Bail Reform Changes Explained."

Thus, even if one is concerned about the obstruction of justice, subpoenas are still a viable alternative to serve as the default. Just as there are some cases in which subpoenas are insufficient in civil cases, there may be some cases in which subpoenas are insufficient in criminal cases. In those cases, arrest may be permissible, as it would fulfill a primary function (i.e., preventing the obstruction of justice) that subpoenas cannot. But those cases are likely to be limited, since there is not a presumption that all criminal defendants would flee, tamper with evidence, or intimidate witnesses. Thus, the obstruction of justice explanation for arrests would only cover a small subset of cases.

3.2 Aiding Justice

Another possible explanation for why arrests are justified, also in the category of ensuring a proper trial, is that the police need to collect information from defendants. A proponent of arrests may argue that the police need to gather basic information from a suspect, including her name and address. Suppose that the police appear on the scene after a crime. They suspect that A has committed the crime, but they do not know A's name or address. Without the name and contact information, it would be difficult to later issue a subpoena, since the police would not know whom to send it to. Thus, perhaps arrests are justified since they appear to fulfill another primary function that subpoenas cannot (i.e., collecting biographical information from suspects).

Arrests, however, are not necessary to find out a person's name and contact information. Instead, a police officer could simply ask for this information. One may be concerned that the person may refuse to share her name and address or that the person may lie about her identity. Yet, in other contexts, the presumption is that people respond truthfully when a police officer asks for their biographical information, even when there is a potential penalty. When a driver is pulled over for a traffic violation, the police do not immediately arrest her. Instead, a police officer asks for the driver's name and for the other information needed.[47] If the police officer then issues a ticket, the person is not required to pay immediately. Instead, the driver is sent a ticket with information about how to pay the associated fine. Similarly, the police officer could ask a suspect for her name and contact information. The police could then later send a subpoena to the address if one was warranted. The concern about identifying the suspect does not require arrests across the board. Rather, arrests would potentially be permissible only when a person refuses to share the necessary information.

[47] In a traffic stop, the police may ask for identification to verify the person's identity. Similarly, the police could ask for identity verification in order to subpoena a suspect in a criminal case.

Second, one may argue that arrests are justified because they fulfill the primary function of allowing the police to question suspects as part of their investigations. The police need to have the suspect in custody in order to conduct an interview.

I have two responses. First, this justification similarly falls to the objections raised above regarding the use of a subpoena as an alternative. If the police need to speak with witnesses or other people who may have information relevant to the investigation, they can request that the person come for an interview. Indeed, people are typically not arrested when the police simply want to question them. Similarly, the police could request, via subpoena or otherwise, to speak with the suspect.

Second, we should be wary of requiring people to speak to the police. Indeed, these "interviews" are often more akin to interrogations. Wealthy, well-informed suspects often do not engage in these conversations and instead invoke their right to remain silent and their right to an attorney. As such, these interviews primarily target poor suspects who are not properly informed of their rights. The interrogations can lead to self-incriminating disclosures or false confessions. Thus, it is often specifically contrary to the suspect's interest to engage in these interviews. Indeed, courts have sought to protect defendants from the potential minefield of these interviews.[48] As a result, ensuring that these interviews take place does not seem like a worthwhile or permissible goal in the first place.

To review, I presented a potential justification for arrests based on the concern that arrests fulfill the primary function of ensuring justice, while subpoenas cannot. I argued that such a justification would only apply in the following, limited subset of cases: (1) the state could demonstrate to a judge that the person is likely to flee or tamper with evidence, (2) the person had already failed to comply with a subpoena, (3) the person refuses to provide her name and contact information to a police officer. Thus, the ensuring justice explanation does not provide grounds for using arrests over subpoenas as the default practice. In the coming section, I explore another type of justification for arrests.

§4. Dangerousness

The second standard justification of arrest is that a person is "dangerous," since she likely committed a crime. It is therefore necessary for the police to arrest her and bring her into custody in order to avert the future threat. This justification comes from the pre-trial detention justification that the practice protects the "safety of any other person and the community."[49]

[48] See, e.g., Miranda v. Arizona, 384 U.S. 436 (1966).
[49] 18 U.S. Code §3142 Release or Detention of a Defendant Pending trial.

Arrests can avert future threats, while subpoenas cannot. Arrests therefore fulfill a primary function that subpoenas cannot, thus satisfying (1). Given that function, the question now is whether the harms of arrest are permissible (2). In the context of pre-trial detention, one of the standard arguments for the practice's permissibility is that a person forfeits some right on the basis of threatening to commit a future crime. If the right against detention is forfeited, the state may permissibly incarcerate the person, since the incarceration harms but does not wrong the person.[50] To justify the harms of arrest, the state could argue that the practice of arrests harms but does not wrong its citizens, since the person forfeited her right against arrest by posing a future threat, thus making the practice permissible.

It is worthwhile noting that, even for pre-trial detention, this explanation does not apply to all criminal defendants. Rather, it is only invoked in cases in which there is evidence that the person would likely commit a crime while on bail. Thus, it is not true that this justification would explain why *all* criminal defendants are arrested. Still, though, one may expect that it could potentially explain why *many* criminal defendants are arrested.

If the justification in those cases is grounded in a future threat, then we need to understand what would be required to justify the harm of arrest on the basis of defense. The justification of arrest and detention would rely on whether the state has good reason to believe that a person poses a threat. If the state has good reason to believe that the person poses a threat, then it seems that the state would be justified in arresting and detaining that person. If the state does not have good reason to believe that the person poses a threat, then it seems that the state would be unjustified in arresting and detaining the person.[51]

The challenge with the dangerousness justification is that the state rarely has proper evidence of a future threat.[52] Let us consider a few different types of evidence. First, suppose that the state has

---

[50] In the pre-trial detention literature, there are two understandings of the rights forfeiture explanation: backward-looking and forward-looking. The backward-looking explanation is that the person forfeited her right by committing a crime. Albert Alschuler presents a version of this argument. See Alschuler, "Preventive Pretrial Detention and the Failure of Interest-Balancing Approaches to Due Process," 556–58. This view has been widely (and correctly) criticized for a variety of reasons, including that detention on these grounds would violate the presumption of innocence. See, e.g., Ashworth, "Four Threats to the Presumption of Innocence"; Myers, "Eroding the Presumption of Innocence"; Baradaran, "Restoring the Presumption of Innocence"; Duff, "Pre-Trial Detention and the Presumption of Innocence." Instead, the primary view in the literature is the forward-looking explanation, namely that a person forfeits her right by threatening to commit a future crime. The arguments presented here are in line with this standard understanding.

[51] The state's justification is evidence-relative, not fact-relative, since the stated justification depends on whether the state possesses good reason to believe that the person poses a threat. If, for instance, the state received flawed information that a person posed a threat, then detention on the basis of this information may be evidence-relative justified, even if it is not fact-relative justified. Thus, the state needs proper evidence in order to justifiably arrest and detain a person.

[52] Frederick Schauer discusses the general issue with finding proper evidence of a future threat such that detention may be warranted. See Schauer, "The Ubiquity of Prevention."

individualized evidence that provides good reason to believe that a person poses a future threat. In such a case, arrest and detention would likely fall into the category of preventive detention. Philosophers have written extensively about what, if anything, justifies preventive detention.[53] If the person clearly poses a threat, then the potential justification of arrest would follow the question of whether or not preventive detention would be justified in a similar case. Engaging with the debate regarding the justification of preventive detention, however, is beyond the scope of the paper.

Individualized evidence, however, is fairly rare. The state often does not and cannot know about the future threat that any given individual may pose. As a result, jurisdictions have turned to risk assessment tools in order to determine whether a person poses a threat.[54] These tools, namely AI-created predictive algorithms, are used to provide a reason for the state to think that the person poses a threat.

A widespread practice of arrest and detention on this basis seems highly problematic, however. First, these tools are notoriously unreliable and racially biased.[55] Often the algorithms were trained with "dirty data"—data that includes patterns of racial bias perpetrated by people.[56] Second, it uses identity-based reference classes to predict the likelihood of a person committing a crime. Thus, it penalizes people on the basis of their membership in a reference class, a practice that has been widely and correctly criticized.[57]

If detention on the basis of these risk assessment tools is problematic, then one could argue that perhaps the state should detain every suspect, since the state is unable to determine who does and does not pose a threat. In this way, the state could avert future threats without relying on AI-created predictive algorithms.

The problem with this approach, however, is that the state would actually still be using statistical evidence. The statistical evidence would be that x% of suspects commit a future crime. Thus, the state would still be penalizing people on the basis of their membership in a reference class. Instead of race or socio-economic status, people would be penalized for their membership in the group

---

[53] See, e.g., Douglas, "Is Preventive Detention Morally Worse than Quarantine?"; Elias, "Rethinking 'Preventive Detention': From a Comparative Perspective: Three Frameworks for Detaining Terrorist Suspects"; Ashworth and Zedner, "Prevention and Criminalization"; Barnett, "Getting Even: Restitution, Preventive Detention, and the Tort/Crime Distinction"; Husak, "Preventive Detention as Punishment?: Some Possible Obstacles"; Zedner, "Preventive Justice or Pre-Punishment?"

[54] See, e.g., Wykstra, "Philosopher's Corner"; Bolinger, "Algorithms and the Individual in Criminal Law."

[55] See, e.g., Hamilton, "The Biased Algorithm."

[56] See, e.g., Richardson, Schultz, and Crawford, "Dirty Data, Bad Predictions." Surprisingly, some states in the US detain pre-trial defendants when risk assessment tools determine that there is only a 16% chance that the defendant would commit an offense between arrest and trial. See Baradaran, "The Presumption of Punishment."

[57] See, e.g., Colyvan, Regan, and Ferson, "Is It a Crime to Belong to a Reference Class."

"suspected of a crime." As such, it faces problems similar to those faced by other forms of statistical evidence that rely on other problematic reference classes.

One may respond that using "suspected of a crime" as a reference group is not problematic, while using race or socio-economic status is. The concern with penalizing people on the basis of membership in a certain group could be that people may not have control over their membership in that group (e.g., race, socio-economic status, age). Membership in the group "suspected of a crime," however, does not have that same feature.

I have three responses. First, people do not have control over their membership in the group "suspected of a crime." True, if a person in fact committed a crime, it is likely that they will be placed into the group of people "suspected of a crime." Still, in almost all of those cases, the person does not opt into that group. Moreover, there are a number of people who have not committed a crime and yet are placed into the group "suspected of a crime." Thus, membership in the group is not within one's control. Second, the group "suspected of a crime" is often a proxy for the group "poor men of color." Thus, it is still discrimination on the basis of race and class. Third, even if a person has control over her membership in a given group, it is still problematic to penalize people on the basis of membership in that group. For instance, people may choose their religion, yet it would be problematic to discriminate on the basis of religion.

One may respond, however, that there is something distinct about the group "suspected of a crime": the label is negative and may be intended to carry censure with it. As a result, treatment that penalizes people for membership in this group may be permissible.

I respond that it is still impermissible to penalize people on the basis of group membership, even if the group seems to call for censure. A famous case in the statistical evidence literature, first introduced by Charles Nesson, illustrates this:[58]

> **Prison Yard**: 100 people are exercising in a prison yard, when 99 of them suddenly join together in a planned attack on the prison guards. The remaining person tries to stop the attack. There is no available evidence distinguishing the innocent person from the rest. Local prosecutors randomly select one person, Polly, and bring her to trial for participating in the attack.

Many people share the strong intuition that it would be problematic to convict Polly solely based on the statistical evidence that 99% of the people in the prison yard participated in the attack. In this case,

[58] The case is adapted from Charles Nesson's case. See Nesson, "Reasonable Doubt and Permissive Inferences."

Polly is a member of the group "convicted of a crime," which is the reason why she is in prison. Indeed, that group seems fairly analogous to the group "suspected of a crime," since membership in both groups (a) is out of one's control, (b) may be a proxy for other identities, and (c) calls for censure. Yet, it seems problematic to convict Polly solely on the basis of her membership in that group, even though 99% of the members of the group committed a second crime. Similarly, it is problematic to arrest and detain people solely on the basis of their membership in the group "suspected of a crime," even if a certain percentage of the group would commit a (future) crime.

Though there are important distinctions between convictions and arrests, this case helpfully demonstrates that the reference class problem still exists, even if the group is "suspected of a crime." Our intuition that there is something different about penalizing people on the basis of their membership in the group "suspected of a crime" likely reflects commonly-held prejudices against criminal defendants. Thus, justifying the practice of arresting every suspect on the basis of statistical evidence is still problematic.

Thus, it appears that it would be difficult for the state to possess proper evidence that a person poses a threat, such that arrest and detention would be justified. Individualized evidence is rare, and statistical evidence—whether AI generated or based on "suspected of a crime" group membership—is both morally and epistemically problematic. In the vast majority of cases, then, there is no proper evidence that criminal defendants pose a threat of committing a future crime. Therefore, without individualized evidence, the state cannot justifiably arrest a suspect on the basis of "dangerousness."

Conclusion

This paper tackled the high-stakes moral issue of arrests, which has received insufficient philosophical attention to date. In the first part of the paper, I developed a challenge for justifying the practice of arrest through the **Subpoena Alternative Argument**. I pointed to a less harmful alternative to ask what justifies the state's choice to use arrests, instead of subpoenas, given the greater harm of arrest. In the second part of the paper, I further developed the challenge by evaluating some likely contenders for potential justifications. Drawing on standard justifications for pre-trial detention, I demonstrated that they would only apply in a limited set of cases. Thus, the pre-trial detention explanations are unable to justify using arrests as the default practice.

The use of arrests instead of subpoenas demonstrates that criminal defendants are presumed to be non-compliant. Just because a person is suspected of committing a crime, the state expects the person to obstruct justice or commit future crimes. In other, non-criminal contexts, there is no

presumption of non-compliance. Rather, the state trusts a person to follow the law, even when doing so will lead to further penalties, and only uses coercive force when there is individualized evidence that the person will not follow the law. For criminal defendants, however, the default is distrust, even though the person is only suspected of, and not convicted of, committing a crime.

The discussion in this paper does not suggest that there are no justifications for the practice of arrest. Rather, it sets up a challenge for justifying the practice and opens the door for further review. Ultimately, if a better justification for the widespread use of arrest is not provided, then the state should switch to using subpoenas as the default practice for bringing criminal defendants to court.

Bibliography

18 U.S. Code §3142 Release or Detention of a Defendant Pending Trial (1984). https://www.law.cornell.edu/uscode/text/18/3142.

Alschuler, Albert W. "Preventive Pretrial Detention and the Failure of Interest-Balancing Approaches to Due Process." *Michigan Law Review* 85, no. 3 (1986): 510–69. https://doi.org/10.2307/1288756.

Alter, Charlotte. "Activists Want to End Police Violence. But Disagree on How to Do It." *Time*, June 5, 2020. https://time.com/5848318/black-lives-matter-activists-tactics/.

Altman, Matthew. *A Theory of Legal Punishment: Deterrence, Retribution, and the Aims of the State*. 1st ed. Routledge, 2021.

Ashworth, Andrew. "Four Threats to the Presumption of Innocence." *The International Journal of Evidence & Proof* 10, no. 4 (July 2006): 241–79. https://doi.org/10.1350/ijep.10.4.241.

Ashworth, Andrew, and Lucia Zedner. "Prevention and Criminalization: Justifications and Limits." *New Criminal Law Review* 15, no. 4 (2012): 542–71. https://doi.org/10.1525/nclr.2012.15.4.542.

Baradaran, Shima. "Restoring the Presumption of Innocence." *Ohio State Law Journal* 72, no. 4 (2011): 723–76.

———. "The Presumption of Punishment." *Criminal Law and Philosophy* 8, no. 2 (2014): 391–406. https://doi.org/10.1007/s11572-013-9236-7.

Barnes, Patricia. "Study Says 'Ban the Box' Policies Hurt (Not Help) Young Minority Male Job Seekers." *Forbes*, sec. Diversity, Equity & Inclusion. Accessed March 5, 2021. https://www.forbes.com/sites/patriciagbarnes/2020/03/01/study-says-ban-the-box-policies-hurt-not-help-young-minority-male-job-seekers/.

Barnett, Randy E. "Getting Even: Restitution, Preventive Detention, and the Tort/Crime Distinction." *Boston University Law Review*, Symposium on the Crime-Tort Distinction, 76 (1996): 13.

Bolinger, Renée Jorgensen. "Algorithms and the Individual in Criminal Law." *Canadian Journal of Philosophy*, Political Philosophy of Data and AI, 52, no. 1 (January 2022): 61–77. https://doi.org/10.1017/can.2021.28.

Bushway, Shawn, Irineo Cabreros, Jessica Welburn Paige, Daniel Schwam, and Jeffrey B. Wenger. "Barred from Employment: More than Half of Unemployed Men in Their 30s Had a Criminal History of Arrest." *Science Advances* 8, no. 7 (2022). https://doi.org/10.1126/sciadv.abj6992.

Calvani, Terry. "Discrimination on the Basis of Arrest Records." *Cornell Law Review* 56, no. 3 (1971).

Clancy, Thomas K. "What Constitutes an Arrest within the Meaning of the Fourth Amendment." *Villanova Law Review* 48 (2003): 129–94.

Colyvan, Mark, Helen M. Regan, and Scott Ferson. "Is It a Crime to Belong to a Reference Class." *Journal of Political Philosophy* 9, no. 2 (2001): 168–81. https://doi.org/10.1111/1467-9760.00123.

Condry, Rachel, and Shona Minson. "Conceptualizing the Effects of Imprisonment on Families: Collateral Consequences, Secondary Punishment, or Symbiotic Harms?" *Theoretical Criminology* 25, no. 4 (2021): 540–58. https://doi.org/10.1177/1362480619897078.

Demby, Gene, and Shereen Marisol Meraji. "A Decade of Watching Black People Die." Code Switch. Accessed June 7, 2023. https://www.npr.org/2020/05/29/865261916/a-decade-of-watching-black-people-die.

Digard, Léon, and Elizabeth Swavola. "Justice Denied: The Harmful and Lasting Effects of Pretrial Detention." New York: Vera Institute of Justice, April 2019. http://www.vera.org/justice-denied-the-harmful-and-lastingeffects-of-pretrial-detention.

Dobbie, Will, Jacob Goldin, and Crystal S. Yang. "The Effects of Pretrial Detention on Conviction, Future Crime, and Employment: Evidence from Randomly Assigned Judges." *American Economic Review* 108, no. 2 (2018): 201–40. https://doi.org/10.1257/aer.20161503.

Douglas, Thomas. "Is Preventive Detention Morally Worse than Quarantine?" In *Predictive Sentencing: Normative and Empirical Perspectives*, edited by Jan W. de Keijser, Julian V. Roberts, and Jesper Ryberg. Wellcome Trust–Funded Monographs and Book Chapters. Oxford, UK: Hart Publishing, 2019. http://www.ncbi.nlm.nih.gov/books/NBK564167/.

Duff, R. A. "Pre-Trial Detention and the Presumption of Innocence." In *Prevention and the Limits of the Criminal Law*, edited by Andrew Ashworth, Lucia Zedner, and Patrick Tomlin, 115–32. Oxford University Press, 2013. https://doi.org/10.1093/acprof:oso/9780199656769.003.0007.

Elias, Stella Burch. "Rethinking 'Preventive Detention': From a Comparative Perspective: Three Frameworks for Detaining Terrorist Suspects." *Columbia Human Rights Law Review*, 2009, 136.

Engel, Robin S., Robert E. Worden, Nicholas Corsaro, Hannah D. McManus, Danielle Reynolds, Hannah Cochran, Gabrielle T. Isaza, and Jennifer Calnon Cherkauskas. "Alternatives to Arrest." In *The Power to Arrest: Lessons from Research*, edited by Robin S. Engel, Robert E. Worden, Nicholas Corsaro, Hannah D. McManus, Danielle Reynolds, Hannah Cochran, Gabrielle T. Isaza, and Jennifer Calnon Cherkauskas, 75–122. Cham: Springer International Publishing, 2019. https://doi.org/10.1007/978-3-030-17054-7_4.

Ferzan, Kimberly Kessler. "Preventive Justice and the Presumption of Innocence." *Criminal Law and Philosophy* 8, no. 2 (June 1, 2014): 505–25. https://doi.org/10.1007/s11572-013-9275-0.

Fines and Fees Justice Center. "A Debtor's Prison: Fines and Fees Create Debilitating Cycle of Incarceration." Accessed October 29, 2022. https://finesandfeesjusticecenter.org/articles/debtors-prison-court-fees/.

Flanders, Chad, Raina Brooks, Jack Compton, and Lyz Riley. "The Puzzling Persistence of Citizen's Arrest Laws and the Need to Revisit Them." *Howard Law Journal* 64 (2020): 161–216.

Foderaro, Lisa W. "New Jersey Alters Its Bail System and Upends Legal Landscape." *The New York Times*, February 7, 2017, sec. New York. https://www.nytimes.com/2017/02/06/nyregion/new-jersey-bail-system.html.

Geffen, Benjamin D. "The Collateral Consequences of Acquittal: Employment Discrimination on the Basis of Arrests without Convictions." *University of Pennsylvania Journal of Law and Social Change* 20, no. 2 (2017): 81–104.

Gershowitz, Adam M. "Justice on the Line: Prosecutorial Screening before Arrest." *University of Illinois Law Review* 2019 (2019): 833–76.

Hamilton, Melissa. "The Biased Algorithm: Evidence of Disparate Impact on Hispanics." *American Criminal Law Review* 56, no. 4 (2019): 1553–78.

Harmon, Rachel. "Why Arrest?" *Michigan Law Review* 115, no. 3 (2016): 307–64. https://doi.org/10.36644/mlr.115.3.why.

Heaton, Paul, Megan Stevenson, and Sandra Mayson. "The Downstream Consequences of Misdemeanor Pretrial Detention." *Stanford Law Review* 69, no. 3 (2017): 711–94. https://doi.org/10.2139/ssrn.2809840.

Hill, Evan, Ainara Tiefenthäler, Christiaan Triebert, Drew Jordan, Haley Willis, and Robin Stein. "How George Floyd Was Killed in Police Custody." *The New York Times*, June 1, 2020, sec. U.S. https://www.nytimes.com/2020/05/31/us/george-floyd-investigation.html.

Hope, Lorraine, Amina Memon, and Peter McGeorge. "Understanding Pretrial Publicity: Predecisional Distortion of Evidence by Mock Jurors." *Journal of Experimental Psychology: Applied* 10 (2004): 111–19. https://doi.org/10.1037/1076-898X.10.2.111.

Hoskins, Zachary. "Collateral Legal Consequences and Criminal Sentencing." *American Philosophical Quarterly* 60, no. 2 (April 1, 2023): 117–30. https://doi.org/10.5406/21521123.60.2.02.

———. "Guilty Pleas, Sentence Reductions, and Non-Punishment of the Innocent." In *Sentencing the Self-Convicted: The Ethics of Pleading Guilty*, edited by Julian V Roberts and Jesper Ryberg, 51–70. Hart Publishing, 2023. https://doi.org/10.5040/9781509957460.

———. "Punishment's Burdens on the Innocent." *Journal of Applied Philosophy*, forthcoming. https://doi.org/10.1111/japp.12642.

Husak, Douglas. "Preventive Detention as Punishment?: Some Possible Obstacles." In *Prevention and the Limits of the Criminal Law*, edited by Andrew Ashworth, Lucia Zedner, and Patrick Tomlin. Oxford University Press, 2013. https://oxford.universitypressscholarship.com/view/10.1093/acprof:oso/9780199656769.001.0001/acprof-9780199656769-chapter-10.

Katz, Matt. "Chokeholds, Chemical Spray, and Head Strikes: Rikers' Federal Monitor Details Abusive Correction Officer Unit." *Gothamist*, April 25, 2023, sec. News. https://gothamist.com/news/chokeholds-chemical-spray-and-head-strikes-rikers-federal-monitor-details-abusive-correction-officer-unit.

Lackey, Jennifer. *Criminal Testimonial Injustice*. Oxford: Oxford University Press, 2023. https://doi.org/10.1093/oso/9780192864109.001.0001.

Leslie, Emily, and Nolan G. Pope. "The Unintended Impact of Pretrial Detention on Case Outcomes: Evidence from New York City Arraignments." *The Journal of Law and Economics* 60, no. 3 (August 2017): 529–57. https://doi.org/10.1086/695285.

Lippke, Richard L. "Punishment Drift: The Spread of Penal Harm and What We Should Do About It." *Criminal Law and Philosophy* 11, no. 4 (2017): 645–59. https://doi.org/10.1007/s11572-016-9392-7.

Menendez, Matthew, Lauren-Brooke Eisen, Michael F. Crowley, and Noah Atchison. "The Steep Costs of Criminal Justice Fees and Fines." Brennan Center for Justice, November 19, 2021. https://www.brennancenter.org/our-work/research-reports/steep-costs-criminal-justice-fees-and-fines.

Merkl, Taryn A. "New York's Upcoming Bail Reform Changes Explained." Brennan Center for Justice, December 10, 2019. https://www.brennancenter.org/our-work/analysis-opinion/new-yorks-upcoming-bail-reform-changes-explained.

Miranda v. Arizona, 384 U.S. 436 (1966) (n.d.).

Myers, Nicole Marie. "Eroding the Presumption of Innocence: Pre-Trial Detention and the Use of Conditional Release on Bail." *The British Journal of Criminology* 57, no. 3 (2017): 664–83. https://doi.org/10.1093/bjc/azw002.

Nesson, Charles R. "Reasonable Doubt and Permissive Inferences: The Value of Complexity." *Harvard Law Review* 92, no. 6 (1979): 1187–1225. https://doi.org/10.2307/1340444.

Neusteter, Rebecca, and Megan O'Toole. "Every Three Seconds: Unlocking Police Data on Arrests." Vera Institute, January 2019. https://www.vera.org/publications/arrest-trends-every-three-seconds-landing/arrest-trends-every-three-seconds/overview.

Ordway, Denise-Marie, John Whibey, and Leighton Walter Kille. "Deaths in Police Custody in the United States: Research Review." Journalist's Resource, June 7, 2020. https://journalistsresource.org/studies/government/criminal-justice/deaths-police-custody-united-states/.

"Paying for Your Time: How Charging Inmates Fees Behind Bars May Violate the Excessive Fines Clause." Brennan Center for Justice. Accessed October 29, 2022. https://www.brennancenter.org/our-work/research-reports/paying-your-time-how-charging-inmates-fees-behind-bars-may-violate.

Richardson, Rashida, Jason M. Schultz, and Kate Crawford. "Dirty Data, Bad Predictions: How Civil Rights Violations Impact Police Data, Predictive Policing Systems, and Justice." *New York University Law Review Online* 94 (2019): 15–55.

Roberts, Anna. "Arrest as Guilt." *Alabama Law Review* 70, no. 4 (2019): 987–1030.

Rosenberg, Carol. "Pentagon's Repatriation of Algerian Leaves 30 Prisoners at Guantánamo." *The New York Times*, April 20, 2023, sec. U.S. https://www.nytimes.com/2023/04/20/us/politics/repatriation-prisoner-guantanamo-bakush.html.

Sainato, Michael. "'Slavery by Any Name Is Wrong': The Push to End Forced Labor in Prisons." *The Guardian*, September 27, 2022, sec. US news. https://www.theguardian.com/us-news/2022/sep/27/slavery-loophole-unpaid-labor-in-prisons.

Scarre, Geoffrey. "Corrective Justice and Reputation." *Journal of Moral Philosophy* 3, no. 3 (January 1, 2006): 305–19. https://doi.org/10.1177/1740468106071223.

Schauer, Frederick. "The Ubiquity of Prevention." In *Prevention and the Limits of the Criminal Law*, edited by Andrew Ashworth, Lucia Zedner, and Patrick Tomlin, 0. Oxford University Press, 2013. https://doi.org/10.1093/acprof:oso/9780199656769.003.0002.

Stevenson, Megan T. "Distortion of Justice: How the Inability to Pay Bail Affects Case Outcomes." *The Journal of Law, Economics, and Organization* 34, no. 4 (November 1, 2018): 511–42. https://doi.org/10.1093/jleo/ewy019.

"The State of Bail: A Breakthrough Year for Bail Reform." Vera Institute, 2017. https://www.vera.org/state-of-justice-reform/2017/bail-pretrial.

Valentine, Ashish. "What Is the Citizen's Arrest Law in the Trial over Ahmaud Arbery's Death?" *NPR*, October 26, 2021. https://www.npr.org/2021/10/26/1048398618/what-is-the-citizens-arrest-law-in-the-trial-over-ahmaud-arberys-death.

Weichselbaum, Simone, Lisa Riordan Seville, Emily Siegel, Joseph Neff, and Abbie VanSickle. "Violent Encounters with Police Send Thousands of People to the ER Every Year." *The Marshall Project*, June 23, 2021, sec. Feature. https://www.themarshallproject.org/2021/06/23/violent-encounters-with-police-send-thousands-of-people-to-the-er-every-year.

Wykstra, Stephanie. "Philosopher's Corner: What Is 'Fair'? Algorithms in Criminal Justice." *Issues in Science and Technology* (blog), April 30, 2018. https://issues.org/perspective-philosophers-corner-what-is-fair-algorithms-in-criminal-justice/.

Zedner, Lucia. "Preventive Justice or Pre-Punishment? The Case of Control Orders." *Current Legal Problems* 60, no. 1 (2007): 174–203. https://doi.org/10.1093/clp/60.1.174.

Pre-Trial Punishment: Demystifying Credit for Time Served

Is pre-trial detention permissible? To answer this question, I begin with two philosophical puzzles that arise from the practice of crediting time served. I contend that the practice of crediting time served provides a presumptive argument in favor of understanding pre-trial detention as punishment, and the two puzzles can be resolved with this understanding. I then argue that pretrial detention fits standard definitions of punishment. Since pre-trial punishment occurs before a person has been convicted, let alone tried, the practice is straightforwardly impermissible. I end by proposing a novel compensation and credit model for pre-trial punishment.

Introduction

In 2010, Kalief Browder, a 17-year-old, was detained at Rikers, a New York City jail, for allegedly stealing a backpack. After over 1,000 days, 700 of which were spent in solitary confinement, Browder was released from the jail and sent home. The charges against him had been dropped. Soon after, he died by suicide.[59]

Browder's tragic death directed public attention to a common, widespread practice in legal jurisdictions: pre-trial detention. Almost half a million people are currently detained pre-trial in the United States.[60] Worldwide, approximately three million people are currently incarcerated pre-trial, comprising roughly one third of the world's incarcerated population.[61]

The time spent in pre-trial detention is often credited towards a future sentence. For example, Dragan Jokić was convicted of war crimes by the International Criminal Tribunal for the former Yugoslavia in 2007. Although he received a nine-year sentence, he had already been in pre-trial detention since 2001. The almost seven years he had spent in pre-trial detention served as a credit towards his future sentence. Jokić's release date was therefore set for 2009, only two years later.[62]

While the practice of crediting time served in pre-trial detention is quite common, it creates a philosophical puzzle. Because people who are detained before trial and then convicted receive credit towards their post-trial sentence, it appears that the time spent in pre-trial detention is considered part of the overall punishment. Yet, scholars across the board say that pre-trial detention is not

[59] Gonnerman, "Kalief Browder, 1993–2015."
[60] Wagner and Sawyer, "Mass Incarceration."
[61] Bisogno et al., "Data Matters."
[62] Blagojević & Jokić (IT-02-60).

punishment.[63] To add to the puzzle, people who are detained pre-trial who are not convicted, like Browder, receive no credit or compensation for that very same time spent in pre-trial detention.[64]

Though pre-trial detention has sustained a fair amount of philosophical critique, the practice of crediting time served has received far less attention. To date, only three articles have given it serious attention.[65] Two of the authors accept the practice of providing credit for time served, while one argues in favor of abolishing crediting time served and replacing it with a compensation scheme.[66] All three of the scholars, however, maintain that pre-trial detention is not punishment.[67]

In this paper, I argue that pre-trial detention is pre-trial punishment—full stop. Though some philosophers have tentatively mentioned this as a possibility or have called it by similar names (e.g., "harsh treatment," "akin to punishment"), I contend that it is punishment (descriptive) and should be recognized as such (prescriptive).[68]

Why does it matter whether pre-trial detention is labeled as punishment? Ultimately, the most important issue is the permissibility of the practice. In order to answer the question of permissibility, however, we must first address the question of whether or not the practice is punishment, since state punishment requires certain unique justifications and procedures for it to be permissible. Thus, if I am correct that pre-trial detention is pre-trial punishment, then the standard justifications and procedures required for permissible punishment would apply to pre-trial detention as well. If, for

---

[63] See, e.g., Alschuler, "Preventive Pretrial Detention and the Failure of Interest-Balancing Approaches to Due Process"; Donelson, "Natural Punishment"; Baradaran, "Restoring the Presumption of Innocence"; Baradaran, "The Presumption of Punishment"; Lippke, "Preventive Pre-Trial Detention without Punishment"; Brunt and Bowman, "Toward a Just Model of Pretrial Release: A History of Bail Reform and a Prescription for What's Next"; Duff, "Pre-Trial Detention and the Presumption of Innocence"; Ferzan, "Preventive Justice and the Presumption of Innocence"; Klein, "The Pretrial Detention 'Crisis'"; Makar, "Unnecessary Incarceration"; Mayson, "Dangerous Defendants"; Stevenson and Mayson, "Pretrial Detention and the Value of Liberty"; Manns, "Liberty Takings: A Framework for Compensating Pretrial Detainees."

[64] The lack of conviction occurs when a jury acquits a defendant or when prosecutors drop the charges against the defendant.

[65] Kolber, "Against Proportional Punishment"; Donelson, "Natural Punishment"; Ferzan, "The Trouble with Time-Served." See Ferzan (p. 5): "Only two articles have given it any serious attention," referring to Kolber and Donelson on crediting time served.

[66] Kolber and Donelson support the practice of crediting time served, while Ferzan contends that it should be fully replaced by compensation.

[67] Kolber argues that it is not punishment. Donelson calls it "natural punishment" (a term he develops for bad consequences that arise from committing a crime) but explains clearly that it is distinct from standard punishment. Ferzan acknowledges that pre-trial detention looks like punishment, but ultimately contends that it should be conceived of as taking a person's time, akin to sequestering jurors.

[68] Kolber refers to pre-trial detention as "harsh treatment," and Makar calls pre-trial detention "akin to punishment." See Kolber, "Against Proportional Punishment"; Makar, "Unnecessary Incarceration," 628. For other descriptions of pre-trial detention, see: "looks a lot like punishment" (Stevenson and Mayson, "Pretrial Detention and the Value of Liberty," 762.), "informal punishment" (Manns, "Liberty Takings: A Framework for Compensating Pretrial Detainees.").

instance, a conviction is necessary for permissible punishment, then we can clearly and easily see that pre-trial detention is impermissible.

In §1, I lay out the two puzzles that arise from crediting time served. In §2, I argue that the practice of crediting time served provides a strong presumptive argument in favor of understanding pre-trial detention as punishment. In §§3–4, I demonstrate that pre-trial detention fits standard definitions of punishment. In §5, I discuss the permissibility of the practice. Finally, in §6, I propose a novel compensation and credit model for pre-trial punishment.

§1. The Puzzles of Credit for Time Served

Before we begin, it is important to clarify two aspects of pre-trial detention. First, pre-trial detention refers to the incarceration of criminal defendants who have been charged with, but not convicted of, a crime. Pre-trial detention therefore typically spans from arraignment, when the person is charged, until a verdict declaration, when a person is convicted or acquitted.[69]

Second, the time spent in pre-trial detention, which is referred to as "time served," is commonly credited towards the post-trial sentence. In the US, all 50 states and the federal government reduce post-trial sentences based on the amount of time spent in pre-trial detention.[70] In the UK, the time spent incarcerated pre-trial is automatically counted towards the post-trial sentence.[71] The practice of crediting time served is the central focus of this paper.

There are two ways to formulate the puzzles that arise from credit for time served: (1) the co-defendants puzzle, and (2) the acquitted defendant puzzle. In this section, I explain these two puzzles.

1.1 The Co-Defendants Puzzle

The formulation of the first puzzle, which I term the co-defendants puzzle, was originally presented by Adam Kolber.[72] Kolber argues that there is an inherent proportionality concern that arises from the practice of crediting time served. Though proportionality is widely debated by scholars, for the purposes of this puzzle, it can be understood simply as the notion that two people who commit the same crime and are equally blameworthy should receive the same punishment.[73]

---

[69] Post-verdict, pre-sentencing detention is beyond the scope of this paper, since it is not pre-trial.
[70] For a list of every relevant statute, see Ferzan, "The Trouble with Time-Served," 4 n. 11.
[71] Criminal Justice Act 2003 s.240ZA Credit for Time Spent on Remand.
[72] Kolber, "Against Proportional Punishment."
[73] Though philosophers disagree about whether proportionality should track harm caused, blameworthiness, or some other feature, many agree that this principle should determine appropriate sentence lengths. For this understanding, Kolber cites

To demonstrate the proportionality challenge that arises from crediting time served, consider the following case:

> **Co-Defendants**: A and B commit the same crime, and both face a 10-year sentence. Both are given the option to await trial at home if they can post bail. While A can post bail, B cannot. A spends two years awaiting trial at home, and B spends those two years in jail. At the trial, both A and B are found guilty and sentenced to 10 years in prison. A spends the next 10 years in prison. B, however, is given a two-year credit for the time she already spent incarcerated. B, therefore, only spends the next eight years in prison.

This scenario creates a puzzle for proponents of proportionality. If A and B commit the same crime and are equally blameworthy, then the proportional punishment principle dictates that they should receive the same punishment—10 years in prison. If the time B spent in pre-trial detention is not punishment, per the standard understanding, then it appears that B only receives an eight-year punishment. B is therefore under-punished relative to both the warranted 10-year sentence and A's 10 years of post-trial incarceration. Consider the following chart:

	Pre-trial	Post-trial	Standard Sentence	Standard – Post-trial
A	0	10	10	0
B	2	8	10	2

Thus, it appears that the practice of crediting time served leads to the disproportional under-punishment of those who are detained pre-trial. If someone is credited for her time served, then the post-trial sentence is shorter than the standard sentence. If the time spent in pre-trial detention had not been credited towards the post-trial sentence, then A and B would have both spent 10 years in prison post-trial—the proportional punishment.

The following three principles conflict to create the **Co-Defendants Puzzle**:

(1) Convicted detainees should receive credit for time served.

(2) Punishment should be proportional.

(3) Pre-trial detention is not punishment.

Given that these three principles all appear to be true, one must deny at least one of the above principles to resolve the **Co-Defendants Puzzle**.

---

a variety of scholars, including Ristroph, "How (Not) to Think Like a Punisher," 738–39; Duff, *Punishment, Communication, and Community*, 135; Husak, "Already Punished Enough," 83.

Let us first consider whether we could deny (1), which states that convicted detainees should receive credit for time served. The rejection of (1) would mean that those who are detained pre-trial should not receive any credit for this time. To understand what denying credit for time served entails, let us consider a revised version of the **Co-Defendants** case in which everything is the same, except that the legal system does not credit time served:

> **Co-Defendants***: B spends two years incarcerated pre-trial and then also serves a full 10-year sentence post-trial. A serves a 10-year sentence post-trial.

In **Co-Defendants***, A and B serve post-trial sentences of the same length. If the two years that B spends incarcerated pre-trial are not punishment, then A and B both receive a 10-year punishment. The equal sentences conform to the proportional punishment principle, as they both receive the proportional punishment, 10 years, for their crime.

Yet, it seems unfair to deny credit for time served. If B is not provided credit for time served, then B ends up spending two years in pre-trial detention and another 10 years in post-trial detention. In total, B spends twelve years incarcerated, while A only spends 10 years incarcerated. Moreover, this route defies the standard practice of crediting time served. Many legal systems around the world provide credit for time served.[74] Thus, denying the practice of crediting time served is not an appealing option for solving the proportionality puzzle.[75]

A second option is to deny or modify (2), which states that punishment should be proportional. Kolber ultimately opts to deny (2) as part of his broader critique of proportionality, which is the central aim of his argument. Alternatively, one could modify (2) to try to resolve the puzzle. Both options, however, are not good routes for resolving the puzzle.

Denying that punishment should be proportional has widespread implications for the general justification and practice of punishment. Proportionality is key to explaining why a person should receive a lighter sentence for petty theft and a harsher sentence for murder. Moreover, proportionality explains why people should be treated consistently across a system of punishment. For instance, if two people commit the same crime and are equally blameworthy, then proportionality would dictate that they should receive the same punishment. Thus, the proportionality principle corresponds to

---

[74] See, e.g., "Credit for Time Spent on Remand"; Blagojević & Jokić (IT-02-60); Ferzan, "The Trouble with Time-Served," 4 n. 11.

[75] Kolber adds a further argument against denying (1), namely that no scholars had actually adopted or advocate for this position. Kolber, "Against Proportional Punishment," 1149. Approximately a decade later, Ferzan does in fact advocate for abolishing credit for time served (along with a dramatic re-imagination of pre-trial detention). See Ferzan, "The Trouble with Time-Served."

basic intuitions that motivate much of the general understanding of fair punishment. For this reason, it appears to be a heavy cost to deny proportional punishment in order to save credit for time served.

Importantly, if we did deny proportionality, it is unclear why one would credit time served in the first place. Consider the scenario in **Co-Defendants***, in which A is not incarcerated pre-trial, and B spends two years incarcerated pre-trial. It seems that a proponent of denying proportionality would have no objection to B serving twelve years of incarceration while A only serves 10 years of incarceration. If we do not care about proportionality, there is no need to make their overall incarcerations consistent with one another. Nor is there a need to make their post-trial incarcerations consistent or to ensure that a given individual serves the correct sentence.

Just as denying (2) is not an appealing route, modifying (2) also faces a set of challenges. Kolber suggests a modified understanding of proportionality, although he ultimately rejects it. Instead of focusing on whether the *punishment* is proportional, he suggests that perhaps we should determine whether the *harsh treatment* is proportional.[76] If the relevant measurement is harsh treatment, then it does not matter whether the harsh treatment was inflicted before or after trial. Thus, A and B are proportionally harmed if B is credited for time served, since they both suffer for a total of 10 years. If proportionality refers to harsh treatment, the puzzle seems resolved.

The proportional harsh treatment view, however, has two types of issues. First, there are a variety of general problems beyond this puzzle with using harsh treatment as the measure of proportionality, making it an unappealing view to adopt. Second, harsh treatment proportionality does not actually do any better than the proportional punishment view, since it too fails to explain certain instances of crediting time served. Instead, the puzzle remains with either understanding of proportionality.

For a variety of reasons, the proportional harsh treatment view is not a good replacement for proportional punishment. Most prominently, people suffer differently from punishment. Different people's psychological dispositions contribute to different experiences of harsh treatment.[77] For instance, a person who is claustrophobic will suffer from incarceration significantly more than a person who is not bothered by tight spaces. If one adopts the harsh treatment view, then the person who is claustrophobic would need to receive a shortened sentence. This view, however, has the perverse upshot that wealthy people would likely be incarcerated for shorter periods of time simply

[76] Kolber, "Against Proportional Punishment." Kolber brings up this alternative understanding of proportionality on p. 1155 and then spends the remainder of the argument against harsh treatment proportionality.
[77] For a more robust argument, see Kolber, "The Subjectivist Critique of Proportionality"; Kolber, "Against Proportional Punishment"; Kolber, "The Comparative Nature of Punishment."

because they would likely find the experience of incarceration especially uncomfortable.[78] Moreover, it is nearly impossible, if not impossible, to tailor the experience of incarceration in this way.

One may respond that proportional harsh treatment does not require individually tailored experiences but rather consistent harsh treatment by the state. But the state does not, and maybe cannot, treat people consistently. Prison conditions, such as the sizes of cells, the number of people incarcerated per cell, the quality of the food, and the programming opportunities vary dramatically across facilities and even within a given facility.

Even if the state could treat people consistently, there is a more fundamental concern with using harsh treatment proportionality to resolve this puzzle. If harsh treatment—not punishment—is what we care about, then a different problem arises with the practice of crediting time served, which I call the **Proportionality Dilemma**. Consider the following case:

> **Co-Detainees**: B and C are both placed in pre-trial detention. After two years, B is convicted of a crime and faces a 10-year sentence. Because of the credit for the two years spent in pre-trial detention, B spends the next eight years incarcerated. C, however, is acquitted and goes home without compensation or any other form of credit.

In this case, B deserves a 10-year sentence, while C deserves a zero-year sentence. Since B is granted credit for time served, B ultimately receives 10 years of incarceration. Thus, the credit for time served makes it seem like B ends up at her deserved well-being baseline. C, however, ultimately receives two years of incarceration. Thus, C is worse off relative to her deserved well-being baseline.[79]

To demonstrate the **Proportionality Dilemma**, let us bring together the **Co-Defendants** case (A & B) and the **Co-Detainees** case (B & C). If we maintain the practice of crediting time served, then A, B, and C receive the following treatment:

	Deserved	Pre-trial	Post-trial	Total
A	10	0	10	10
B	10	2	8	10
C	0	2	0	2

[78] As Doug Husak notes, "Few suggestions are more distasteful to the public than that the privileged, in virtue of their elevated status, should be punished less severely than the disadvantaged." Husak, "Already Punished Enough," 82.

[79] Ferzan is one of the first in the literature to point out the surprising notion that an acquitted defendant is worse-off relative to a convicted defendant. This leads her to call crediting time served a "perverse" practice that "harms the innocent." She does not, however, see it as part of a puzzle. See Ferzan, "The Trouble with Time-Served."

This chart reflects how sentencing would work across jurisdictions. If pre-trial detention is not punishment, then neither understanding of proportionality can fully explain the chart above. Each understanding of proportionality helps explain one case but cannot explain the other.

Suppose that proportionality is proportional punishment. If so, then the only relevant comparison is between the "Deserved" and the "Post-trial" columns, since pre-trial detention is not punishment and is therefore irrelevant. A and C serve the deserved post-trial sentences. Indeed, this understanding of proportionality explains C's treatment in the **Co-Detainees** case—C is not credited or compensated since C is never punished. Yet, as we saw in the **Co-Defendants** case, B's shortened sentence of eight years is disproportional relative to both the deserved 10-year sentence and A's 10 years of post-trial punishment. Thus, the proportional punishment understanding is a problem for the **Co-Defendants** case.

Alternatively, suppose that proportionality is proportional harsh treatment. If so, then the only relevant comparison is between the "Deserved" and the "Total" columns, since distinguishing between pre-trial and post-trial harsh treatment is irrelevant. A and B experience the same harsh treatment (more or less), since they both deserve and experience a total of 10 years of incarceration. This understanding of proportionality therefore explains the treatment of A and B in the **Co-Defendants** case. Yet, as we saw in the **Co-Detainees** case, C's two years of incarceration pose a problem. C receives two years of harsh treatment, which is the relevant measure for A and B's treatments. Yet, C deserves zero years of harsh treatment, so C receives disproportional harsh treatment. Thus, the proportional harsh treatment understanding is a problem for the **Co-Detainees** case.

Thus, the dilemma is as follows: proportional punishment can help explain the **Co-Detainees** case but fails to explain the **Co-Defendants** case, while proportional harsh treatment can help explain the **Co-Defendants** case but cannot explain the **Co-Detainees** case. So, neither modifying nor denying (2) is a good option for resolving the puzzle.

Given that both understandings of proportionality cannot explain the practice of crediting time served, the best way out of the dilemma may be to accept the asymmetry between pre-trial detention of a person who is later convicted and pre-trial detention of a person who is later acquitted. This leads us to our second puzzle.

1.2 Acquitted Defendant Puzzle

There appears to be an asymmetry in the treatment of pre-trial detention. If a person is detained pre-trial and then convicted, the time spent in pre-trial detention seems to count as punishment. If a person is detained pre-trial and then acquitted, that very same time does not seem to count as punishment and thus the person is not compensated or given credit towards future sentences.

Thus, the **Acquitted Defendant Puzzle** arises from the following three premises:

(1) The pre-trial detention of a person who is later convicted is punishment if and only if the pre-trial detention of a person who is later acquitted is punishment.

(2) Pre-trial detention of a person who is later convicted is punishment.

(3) Pre-trial detention of a person who is later acquitted is not punishment.

Let us first consider whether we can deny (1). Raff Donelson presents arguments that can be used to deny (1). Though he does not consider the differential treatment in the context of an acquitted defendant, he does note that there is a double treatment of pre-trial detention.[80] Before conviction, pre-trial detention is not considered punishment. Yet, after conviction, it is retrospectively counted as part of the punishment. It appears that, for Donelson, the conviction turns the past treatment from non-punishment into punishment.

If one applies Donelson's reasoning to an acquitted defendant, we see that pre-trial detention of a later convicted defendant and pre-trial detention of a later acquitted defendant are not the same. When a defendant is convicted, the past treatment is turned from non-punishment into punishment. But since the acquitted defendant is never convicted, the past treatment is never turned from non-punishment into punishment. Thus, the pre-trial detention of a person who is later convicted is different from pre-trial detention of a person who is later acquitted—one has been retroactively deemed punishment, while the other has not.

There are a number of reasons to reject this retroactive argument. First, the treatment appears identical, regardless of whether the person is later convicted or acquitted. Everyone who is detained pre-trial loses the same liberties, and the experience is the same across the board, irrespective of future verdicts. The state provides the same justification for detaining both people who are later convicted and people who are later acquitted.

Second, the state does not provide any sort of explanation as to *why* the pre-trial detention of people who are convicted is retroactively deemed punishment. Other cases of retroactive legal

[80] Donelson, "Natural Punishment."

redesignation are disanalogous in that sense. For instance, when a marriage is annulled, the annulment retroactively makes it such that the marriage never occurred. Yet, there are explanations as to why the marriage was never valid, such as that one partner never consented to the marriage or that one partner was underage at the time of the marriage. Not only do these provide an explanation for the annulment when it happens, they also explain the ways to make a marriage ineligible for annulment at the start of the marriage. In the context of pre-trial detention, however, there is no such explanation given. Similarly, there is no explanation as to how one could make pre-trial detention eligible or ineligible for designation as punishment in the future. Factual guilt and factual innocence cannot play this role, since some factually guilty people are acquitted and not credited, while some factually innocent people are convicted and credited. Moreover, if factual guilt is needed for something to constitute punishment, then the incarceration of people who are innocent and wrongfully convicted would not constitute punishment.

Finally, the retroactive argument relies on the state's future declaration that something is punishment. If the state's declaration (e.g., a conviction) is needed for retroactively designating a treatment as punishment or not punishment, then whether pre-trial detention is punishment rests exclusively on the state. It is highly dubious that the state has the power to designate the same treatment as punishment or not punishment in accordance with its own interests. When a person is acquitted, the state calls the treatment not punishment so that the state does not have any compensatory duties. When a person is convicted, the state calls the treatment punishment and reduces the post-trial sentence, thus making it cheaper to carry out the punishment. More generally, requiring the state to call something "punishment" in order for a treatment to constitute punishment is problematic for a number of reasons that I address later in the paper.

For all of the above reasons, the retroactive explanation as to why the pre-trial detention of a person later convicted is different from the pre-trial detention of a person later acquitted is problematic. Thus, denying (1) is not a good solution to the puzzle.[81]

Indeed, scholars such as Kolber, Donelson, and Kimberley Ferzan would choose to deny (2), which states that credit for time served demonstrates that the pre-trial detention of a person who is later convicted is punishment.[82] (Indeed, the denial of (2) leaves one with no explanation for the

[81] Kolber also presents a brief argument against the notion of retroactive redesignation, in part because "most people likely think that whether conduct is punishment depends on the conduct at the time it occurs and not on subsequent events." Kolber, "Against Proportional Punishment," 1151.

[82] Ferzan, "The Trouble with Time-Served"; Kolber, "Against Proportional Punishment"; Donelson, "Natural Punishment." All future references to these scholars are from these articles unless indicated otherwise.

practice of crediting time served, which is precisely what puzzles these scholars.) Similarly, these scholars would accept (3), which states that the lack of credit or compensation demonstrates that the pre-trial detention of a person who is later acquitted is not punishment. Whether pre-trial detention is punishment (2) or is not punishment (3) is our central question.

§2. Demystifying Credit for Time Served

Contrary to the three scholars who have written on credit for time served, I argue that pre-trial detention is punishment. The practice of crediting time served provides a presumptive argument for the notion that pre-trial detention is punishment.

First, all jurisdictions in the US, including federal courts, are required to provide credit towards post-incarceration based on the amount of time spent in pre-trial detention.[83] Thus, it is widely treated as part of the overall punishment. Second, the time is credited in a 1:1 ratio; one year spent incarcerated pre-trial is equivalent to one year spent incarcerated post-trial. Furthermore, pre-trial detention is *only* credited towards the sentence for the crime for which the person was detained pre-trial in the first place. It is not a general "get out of prison early" credit. Thus, it appears to be punishment specifically for the crime for which the person is incarcerated.

The practice of crediting time served is not much of a mystery if pre-trial detention is punishment. Simply put, the punishment starts when the incarceration starts. Indeed, Scotland calls the practice of crediting time served "backdating the sentence," which describes the practice exactly as it is.[84] Both of the puzzles can be resolved if we understand that pre-trial detention is punishment.

If pre-trial detention is punishment, which entails denying (3) of the **Co-Defendants Puzzle**, we see that B is not under-punished. B is punished for two years, tried, and then punished for eight more years. Altogether, B receives 10 years of punishment. B and A, therefore, are punished equally—B just begins the punishment two years before A does.

We can also resolve the **Acquitted Defendant Puzzle** if pre-trial detention is punishment, namely by denying (3). If pre-trial detention is punishment, we can better understand why convicted defendants are credited, while acquitted defendants do not receive anything, without resorting to a retrospective designation argument. The time spent in pre-trial detention is punishment specifically for the crime for which the person is detained. Thus, the credit only works towards the sentence for that specific crime. Acquitted defendants are not given a credit towards future sentences likely for the

[83] Ferzan, "The Trouble with Time-Served," 4 n. 11.
[84] Criminal Procedure (Scotland) Act of 1995, c. 46, § 210, also cited in Ferzan, 5 n. 12.

same reasons that we do not give punishment vouchers to innocent people who are exonerated (e.g., it incentivizes bad behavior and weakens the expressive link between a crime and its punishment).

One may expect, then, that acquitted defendants would receive compensation, since innocent people who are exonerated are often compensated. Unfortunately, the practice of not compensating acquitted pre-trial detainees is consistent with the widespread practice of not compensating innocent people who are exonerated. In over 21 states in the US, there is no law on the books requiring the state to compensate people who are exonerated.[85] Thus, the lack of compensation for acquitted pre-trial detainees is sadly consistent with the lack of compensation for innocent people who are punished post-trial and then exonerated. Thus, if we understand pre-trial detention as punishment, then the **Acquitted Defendant Puzzle** is not puzzling; it is just another instance of the unfortunate practice of not compensating innocent people who are punished.

In addition to solving the puzzles, understanding pre-trial detention as punishment is also more consistent with public conceptions of pre-trial detention. In the US, for instance, pre-trial detention happens in jails, while post-trial detention happens in prisons. Yet, many people, including journalists, scholars, and even attorneys, use the terms "jail" and "prison" interchangeably, in part because people do not even know that there is a distinction between the two.[86] Furthermore, the conditions of pre-trial detention are often very similar to the conditions of post-trial detention, if not worse.[87] US jails are notorious for their terrible conditions.[88]

Many scholars who have written on pre-trial detention acknowledge that pre-trial detention looks a lot like punishment.[89] Yet, to my knowledge, across the board scholars stop short of calling it punishment. In the coming sections, I provide further support for the notion that pre-trial detention is punishment.

§3. Hart's Definition of Punishment

Before we evaluate whether pre-trial detention is punishment, we need to clarify what precisely we are referring to when we say pre-trial detention. The distinction between the current practice of pre-trial

[85] Slifer, "How the Wrongfully Convicted Are Compensated for Years Lost."
[86] Burns, "Assessing Jail Coverage in Introductory Criminal Justice Textbooks," 94. This source is also cited in Ferzan, "The Trouble with Time-Served," 4 n. 9.
[87] The fact that jails often have worse conditions than prisons has been true for many decades. See, e.g., Miller and Guggenheim, "Pretrial Detention and Punishment," 369, citing an article published in 1974.
[88] Vansickle and Villa, "Who Begs to Go to Prison? California Jail Inmates."
[89] See, e.g., Duff, "Pre-Trial Detention and the Presumption of Innocence"; Lippke, "Preventive Pre-Trial Detention without Punishment"; Heard, "Pre-Trial Detention and Its Overuse"; Mayson, "Dangerous Defendants"; Myers, "Eroding the Presumption of Innocence."

detention and an imagined form of pre-trial detention is obscured in the literature. Many scholars have acknowledged that the current practice of pre-trial detention is highly problematic.[90] Philosophers have also criticized the terrible conditions of pre-trial detention, especially those in US jails, or the over-inclusive nature of the current practice.[91]

In the literature, philosophers often oscillate between talking about the reality of pre-trial detention and theoretical versions of pre-trial detention—without clearly distinguishing between the two. This leads to a confusing discourse, since pre-trial detention in theory and in practice can be vastly different. On the one hand, in our current practice, there are instances of cruel conditions with very little justification for detention.[92] People may be held in US jails for years at a time, in conditions so terrible that people opt for plea deals just to be able to leave the jail and go to a prison instead.[93] Moreover, people may be detained only because they are too poor to post bail, without even posing a clear flight or danger risk.[94] On the other hand, per Ferzan's suggestion, we can imagine a theoretical scenario in which defendants are detained in a similar manner to the way in which juries are sequestered during trial proceedings.[95] These dramatically different situations should not necessarily be placed under the same umbrella; one could be permissible and the other impermissible. There is no reason that we should look for the two to have the same justification—each deserves its own analysis. For this paper, I focus exclusively on the current practice of pre-trial detention, both because it is the most pressing issue and because it is not feasible to engage with every theoretical version of pre-trial detention. So, is the current practice of pre-trial detention punishment?

---

[90] See, e.g., Sifferd and Fagan, "Pretrial Detention and Moral Agency"; Brunt and Bowman, "Toward a Just Model of Pretrial Release: A History of Bail Reform and a Prescription for What's Next"; Klein, "The Pretrial Detention 'Crisis'"; Heard, "Pre-Trial Detention and Its Overuse"; Duff, "Pre-Trial Detention and the Presumption of Innocence"; Myers, "Eroding the Presumption of Innocence"; Lippke, "Preventive Pre-Trial Detention without Punishment."

[91] See, e.g., Ashworth, "Four Threats to the Presumption of Innocence"; Duff, "Pre-Trial Detention and the Presumption of Innocence"; Mayson, "Dangerous Defendants"; Stevenson and Mayson, "Pretrial Detention and the Value of Liberty"; Baradaran, "The Presumption of Punishment"; Baradaran, "Restoring the Presumption of Innocence"; Ferzan, "Beyond Crime and Commitment."

[92] See, e.g., Gonnerman, "Kalief Browder, 1993–2015."

[93] Vansickle and Villa, "Who Begs to Go to Prison? California Jail Inmates."

[94] Fellner, "The Price of Freedom."

[95] Ferzan, "The Trouble with Time-Served." For other suggested models of pre-trial detention, see Miller and Guggenheim, "Pretrial Detention and Punishment"; Brunt and Bowman, "Toward a Just Model of Pretrial Release: A History of Bail Reform and a Prescription for What's Next"; Stevenson and Mayson, "Pretrial Detention and the Value of Liberty"; Wiseman, "Pretrial Detention and the Right to Be Monitored."

3.1 The Need for a Neutral Definition

Let us start with H. L. A. Hart's definition of punishment, which is the most commonly cited and widely accepted definition.[96] Hart's definition is intentionally justification-neutral.[97] Before Hart, many definitions of punishment included elements of what justifies punishment.[98] These definitions create what Hart terms a "definitional stop," meaning that cross-theory discussion is difficult or non-existent so long as no one can agree on what is punishment.[99]

Indeed, even in the literature today, the distinction between a definition of punishment and a definition of justified punishment is often blurred.[100] A definition of punishment should cover both instances of justified punishment and instances of unjustified punishment. A definition of justified punishment should exclude instances of unjustified punishment and would likely include conditions that pertain to the justification of punishment. This leads to as many different definitions of justified punishment as there are theories of punishment. Take the two primary types of justifications of punishment—consequentialist and retributivist. For a direct consequentialist, a definition of justified punishment would include, as a necessary condition, that the punishment maximizes utility. For a retributivist, a definition of justified punishment would include, as a necessary condition, that the person being punished deserves the punishment.

A definition of justified punishment can also aim to capture legally justified punishments. For instance, proper conviction is typically a necessary but not sufficient condition for justified punishment. If the state is aware that the person is innocent, then the continued punishment of that person is not justified, even if there was a proper conviction. Thus, the state's well-founded belief in the person's guilt may also be a necessary condition for justified punishment. Different jurisdictions, however, have different definitions of legally justified punishment. For instance, in some jurisdictions, the definition may include the condition that the punishment is consistent with the sentencing scheme

---

[96] For the definition, see Hart, *Punishment and Responsibility*, 4–5. According to Google Scholar, this essay has been cited over 5,500 times (as of May 2023). For a few examples of discussion or use of Hart's definition, see Pulman, *Hart on Responsibility*; Husak, "A Framework for Punishment"; McPherson, "Punishment"; Bedau, "Retribution and the Theory of Punishment"; Feinberg, "The Expressive Function of Punishment"; Lacey, *State Punishment*; Raz, "H. L. A. Hart (1907–1992)"; Burgh, "Do the Guilty Deserve Punishment?"; Smilansky, "The Time to Punish"; Sverdlik, "Punishment."

[97] Hart's definition is closely related to the definitions of Anthony Flew and S. I. Benn. See Flew, "The Justification of Punishment"; Benn, "An Approach to the Problems of Punishment." After *Punishment and Responsibility* was published, the definition was often referred to as the Flew-Benn-Hart definition. In the contemporary literature, however, the definition is primarily associated with Hart.

[98] This problem was pointed out in 1961 (before the publication of *Punishment and Responsibility*) by K.J. Armstrong: "Clearly the logical order is first to decide what punishment is, then to decide whether this thing is morally justifiable or not." Armstrong, "The Retributivist Hits Back," 476.

[99] Hart, *Punishment and Responsibility*, 5.

[100] For instance, Donelson argues that "punishment" must refer to legal punishment. His argument, then, faces this precise issue. See Donelson, "Cruel and Unusual What? Toward a Unified Definition of Punishment."

at the time of conviction, while in other jurisdictions the definition may instead require that the punishment be consistent with the current sentencing scheme. Thus, a punishment that is consistent with an original conviction but inconsistent with a new sentencing scheme would be considered legally justified punishment by some and legally unjustified punishment by others. This is by no means an exhaustive discussion. Rather, it simply demonstrates the myriad of ways in which people can define justified punishment.

In addition to providing an avenue for cross-theory discussion, differentiating between definitions of punishment and definitions of justified punishment has other benefits. If we do not differentiate between the two, the definitions may be improperly over-inclusive or under-inclusive relative to the general understandings of these terms. Moreover, the definitions can serve as tools for evaluating different practices. If we want to know if something is punishment, we need a definition that will include both justified and unjustified punishment. If we want to know if something is a justified punishment, we need a definition that only includes justified punishment. Though this distinction may seem straightforward, the lack of differentiation between punishment and justified punishment has led scholars to the conclusion that something is not punishment only because it is unjustified.[101]

Because of its neutrality, Hart's definition has been widely accepted. Scholars have certainly raised concerns about the definition.[102] Yet, very few people have attempted to provide a new definition of punishment since Hart's.[103] As such, Hart's definition has remained the most standard, commonly cited definition of punishment and thus provides a good starting point.

3.2 Applying Hart's Definition

I contend that pre-trial detention meets the criteria of Hart's definition. Hart's definition of punishment is as follows:

(i) It must involve pain or other consequences normally considered unpleasant.

(ii) It must be for an offence against legal rules.

---

[101] For instance, Donelson argues that pre-trial detention, if punishment, would violate the US Due Process clause that guarantees that a person will not face punishment before trial. He then concludes that it must not be punishment, without exploring the possibility that it could be unconstitutional punishment. Donelson, "Natural Punishment."

[102] For critiques of Hart's definition, see, e.g., Husak, "A Framework for Punishment," 2014; McPherson, "Punishment"; Feinberg, "The Expressive Function of Punishment"; Samek, "Punishment"; Sverdlik, "Punishment."

[103] Yaffe also notices this phenomenon in the literature: "Insofar as the primary question for theorists is the question of justification, then, it is no surprise that philosophers, in the wake of Hart, mostly stopped trying to identify the necessary and sufficient conditions of punishment." Yaffe, "The Norm Shift Theory of Punishment," 481.

(iii) It must be of an actual or supposed offender for his offence.

(iv) It must be intentionally administered by human beings other than the offender.

(v) It must be imposed and administered by an authority constituted by a legal system against which the offence is committed.[104]

Per (i), pre-trial detention entails a loss of liberty and often harsh confinement conditions, which are "consequences normally considered unpleasant." Per (ii) and (iii), pre-trial detention is inflicted after a person has been charged with "an offense against legal rules." Furthermore, pre-trial detainees are "actual or supposed" offenders. If a person did commit the crimes for which she is accused, then she is an actual "offender." If she did not commit the relevant crimes, then she is a supposed "offender," since the state has indicted her, indicating a belief in her likely guilt. Per (iv) and (v), pre-trial detention is "intentionally administered by human beings" with "authority constituted by a legal system against which the offence is committed." When a person is accused of committing a crime in the UK, for example, a UK judge may sentence the person to pre-trial detention. A bailiff, warden, or guard may incarcerate the person in a UK jail. Since pre-trial detention satisfies all five criteria, it constitutes punishment according to Hart's definition.

One may be concerned that pre-trial detention is not properly "for an offence," per criteria (ii) and (iii) of Hart's definition.[105] Indeed, the word "for" has a very broad semantic range.[106] To determine its most likely meaning in this context, consider the semantics of the word "for" in the following three phrases, each of which contains Hart's phrase "for an offence": arrested for an offence, detained for an offence, imprisoned for an offence. In all three of these phrases, the word "for" indicates that the action is done on account of the specific offence. In other words, the word "for" in the phrase "for an offence" likely introduces "a thing, such as a commodity, piece of work, misdemeanor, etc., on account of which a payment, fine, remuneration, etc., is made."[107] With this contextualized definition of "for" in Hart's criteria (ii) and (iii), it is clear that pre-trial detention is "for" an offence. Again, this may be demonstrated by the practice of crediting time served. As discussed previously, credit for time served can only be applied to a sentence for the crime for which one is originally detained pre-trial. Suppose that a person is detained pre-trial for two years for crime x. Suppose further that, soon after the charges for crime x are dropped against her, she is charged with

[104] Hart, *Punishment and Responsibility*, 4–5.
[105] The ambiguity of the word "for" in Hart's definition has been previously noted but not resolved. See, e.g., Spjut, "Criminal Law, Punishment, and Penalties."
[106] See "For, Prep., Conj., n., and Adv." Oxford English Dictionary.
[107] See "For, Prep., Conj., n., and Adv." II, Section 6.a, Oxford English Dictionary.

crime y. In such a scenario, the two years of pre-trial detention would not be credited towards a sentence for crime y. As such, it seems that pre-trial detention is properly "for" an offence and thus fits Hart's definition of punishment.

3.3 Intentionality Objections

One may object that, according to Hart's definition, the treatment needs to be intended as punishment to qualify as such. Indeed, Kolber and others have understood the phrase "intentionally administered" in criterion (iv) to mean that the treatment must be intended as punishment.[108] If so, then perhaps pre-trial detention is not punishment if it is not intended as such.

There are two ways of understanding this objection. The first pertains directly to Hart's definition, and the second pertains to the broader question of if we should care about whether something is intended as punishment when designating it as punishment. I will respond to each understanding of this objection in turn.

First, I disagree with the interpretation that Hart requires the intention to punish in order for harsh treatment to constitute punishment. Hart's phrase "intentionally administered" from the fourth criterion does not refer to the harsh treatment being intended specifically as *punishment*, as opposed to other types of harsh treatment that are not intended as punishment. Rather, it seems that Hart uses this phrase to rule out non-deliberate harms. He writes "intentionally administered" in order to exclude what would typically be referred to as accidental harm or natural harm. Similarly, in the same criterion, he requires that the treatment be inflicted by someone "other than the offender" to exclude self-harm. Hart also stipulates that the treatment be inflicted by a "human being" in order to rule out what is sometimes understood as divine punishment. Taken in the context of the whole criterion, it is clear that the phrase "intentionally administered" does not pose a further requirement that the person inflicting the harm intends the harm specifically as punishment.[109] Rather, the act just needs to be deliberately administered.

Hart's use of the word "intention" in other contexts further supports my understanding of his fourth criterion. Later in his book, Hart argues that the intentionality requirement for criminal convictions refers to the "intention to do the act forbidden by law," as opposed to an intention to

[108] Kolber, "Against Proportional Punishment," 1144. See also, e.g., Berman, "Punishment and Justification"; Miller and Guggenheim, "Pretrial Detention and Punishment," 367; Husak, "Lifting the Cloak: Preventive Detention as Punishment"; Stevenson and Mayson, "Pretrial Detention and the Value of Liberty."
[109] See, e.g., McPherson, "Punishment," 22, who also interprets Hart in this way.

specifically break the law.[110] For example, suppose that a person driving 70 miles per hour is unaware of the 50 miles per hour speed limit. As long as she knowingly drove 70 miles per hour, her act is considered intentional, even if she did not know about the speed limit and thus did not intend to specifically break the law. Thus, "intentional" refers to the deliberate nature of an action, not the aim or purpose of the agent. Thus, an authority can "intentionally administer" harsh treatment that constitutes punishment, regardless of whether the authority aims to specifically punish.

Since there are a variety of reasons to understand Hart's definition as not requiring that the treatment be intended as punishment, it seems that the intentionality objection as it pertains to Hart's definition does not stand. Still, one could argue that whether or not something is intended as punishment should bear on whether the treatment is punishment, regardless of Hart's definition.

This brings us to the second understanding of the intentionality objection. Across the literature, scholars point to the notion that pre-trial detention is not intended as punishment.[111] A key piece of evidence for this argument is the US Supreme Court's decision in United States v. Salerno (1987), which states that pre-trial detention is not punishment.[112] In the decision, the court finds that pre-trial detention is a "regulatory" practice, not a "punitive" one. The court asserts that pre-trial detention is not intended to punish defendants, since it serves other purposes, such as preventing defendants from committing a crime before trial. If pre-trial detention is not intended as punishment, so the argument goes, then it cannot be punishment.

Does the court's declaration that pre-trial detention is not meant to be punishment mean that it is not, in fact, punishment? No. Indeed, there are several reasons to reject the court's statement that pre-trial detention is not intended as punishment.

First, the argument presented in the Salerno opinion—that pre-trial detention is intended to be a regulatory practice—does not actually preclude the possibility that the practice is also a punitive one.[113] Pre-trial detention could both prevent a person from committing future crimes *and* punish the person. In fact, standard post-trial incarceration typically does both. For instance, one of the aims of post-trial incarceration may be incapacitation, stopping the person from committing future crimes.

---

110 Hart, *Punishment and Responsibility*, 116.

111 See, e.g., Kolber, "Against Proportional Punishment," 1133–34; Donelson, "Natural Punishment"; Ferzan, "The Trouble with Time-Served."

112 United States v. Salerno 481 U.S. 739 (1987). All future references to the case are from this decision.

113 This misunderstanding is found across the literature. See, e.g., Raff Donelson's argument: "Criminal suspects are detained, not to exact retribution, but rather to ensure their appearance at trial. That is the intention we must impute to criminal justice officials, unless evidence suggests otherwise." Donelson, "Cruel and Unusual What? Toward a Unified Definition of Punishment," 26.

Yet, it is clear that the presence of this purpose does not negate the possibility that the incarceration is punishment. Both pre-trial and post-trial incarceration can fulfill regulatory aims, such as incapacitation, while also constituting punishment. Thus, the state's description of the purpose of pre-trial detention does not entail that pre-trial detention is not punishment.

Second, the court and its jurisdictions treat pre-trial detention as punishment. As discussed above, the actual practice of crediting time served, endorsed by all jurisdictions in the US, undermines the notion that the time spent in pre-trial detention is not punishment. Furthermore, pre-trial detention not only denies people their liberty but also makes them suffer immensely. If actions speak louder than words, then perhaps we should look to the state's actions (crediting time served, incarcerating people in harsh conditions), instead of the state's words (that it is "regulatory").

Third, a state labeling something as "not punishment" does not entail that the given practice is in fact not punishment. If an authority's classification of something as punishment is necessary for something to be punishment, we would have a bizarre way of defining practices. For instance, if a court asserted that post-trial incarceration is not punishment, then according to this argument, post-trial incarceration would, in fact, not be punishment. It would be odd for the question of whether a treatment is punishment to turn on whether the courts want to call it punishment.

Indeed, there are reasons for the court to be motivated to classify pre-trial detention as "not punishment." If pre-trial detention is punishment, then a whole host of constitutional protections apply, including a right to a fair trial. This would lead to two options: unconstitutional punishment or abolishing pre-trial detention. If a fair trial had to occur before pre-trial detention, then it definitionally would become post-trial detention. Thus, the court may be motivated to call it "not punishment" so that it could maintain the practice of pre-trial detention and ensure that these constitutional rights do not apply. Indeed, Justice Marshall, in his dissent in US v. Salerno, points out the problematic nature of this maneuver:

> The majority's technique for infringing this right is simple: merely redefine any measure which is claimed to be punishment as "regulation," and, magically, the Constitution no longer prohibits its imposition.[114]

In fact, Gideon Yaffe outlines a variety of ways in which the US Supreme Court has classified punitive practices as "not punishment" in order to dodge the constitutional protections afforded to those who are subjected to punishment.[115] According to Yaffe, there is a pattern of the Supreme Court classifying

[114] United States v. Salerno 481 U.S. 739 (1987) at 760.
[115] See Yaffe, "The Norm Shift Theory of Punishment." All future references to Yaffe's view are from this article.

practices that appear to be punitive as "not punishment" in order to shield the practice from the checks required for a practice considered punishment. The court's argument with regard to pre-trial detention can be understood as part of this very pattern.

Thus, even though the court claims that pre-trial detention is not punishment, we have three reasons to reject their argument. First, its argument that pre-trial detention is merely "regulatory" does not entail that it is not also punitive. Second, the practice of crediting time served (and the details of how the credit works) suggests that the court itself often treats it as punishment. Third, labeling something as not punishment to shield the practice from constitutional scrutiny does not mean that the practice is not punishment. Given its context, the decision in US v. Salerno does not appear to be a sufficient reason to conclude that pre-trial detention is not punishment.

Because of this concern about labeling instances of punishment as "not punishment," Yaffe develops a new definition of punishment that specifically does not rely on the state's declaration that a practice is punishment. His definition is the one of the most recently proposed definitions of punishment and one of the few neutral definitions proposed since Hart's definition.[116] In the coming section, I explore whether pre-trial detention meets Yaffe's definition of punishment.

§4. Yaffe's Definition of Punishment

Yaffe distinguishes between punishment-by-intention and punishment-by-nature. Punishments-by-intention are treatments that the state calls punishment, while punishments-by-nature are treatments that are punishments, regardless of the state's stated aim. Since the court states that pre-trial detention is not punishment, pre-trial detention is not punishment-by-intention, according to Yaffe's distinction. That still leaves open the question of whether pre-trial detention is punishment-by-nature.

I argue that pre-trial detention fits Yaffe's proposed definition of punishment-by-nature. Yaffe provides the following definition:

[116] Part of the reason that so few definitions have been proposed is that many scholars think that defining punishment is not necessary, since we only care about whether a practice is justified. See, e.g., McPherson, "Punishment." The definitions that have been proposed tend to be non-neutral, theory-specific definitions of punishment. For instance, many retributivists cite David Boonin's condition that a treatment must be inflicted "*so that* he will suffer" in order for it to constitute punishment. See Boonin, *The Problem of Punishment*, 13. Scholars who subscribe to communicative justifications of punishment cite Joel Feinberg's view that punishment must serve a symbolic function to express condemnation. See Feinberg, "The Expressive Function of Punishment."

> **Norm Shift Theory of Punishment**: P is a punishment-by-nature if and only if (1) P is a required meaningful reduction of a person D's normative freedom and (2) P occurs thanks to government S's exercise of its punitive power. [117]

To demonstrate how pre-trial detention fits this definition, let us consider each part.

First, P must be a required meaningful reduction of a person D's normative freedom. Yaffe begins by distinguishing normative freedom, what one can permissibly do, from metaphysical freedom, what one can physically do. Yaffe clarifies that the term "permissible" does not necessarily refer to moral permissibility. More commonly, it refers to legal permissibility. In the context of pre-trial detention, a person's normative freedom is reduced because it is no longer legally permissible for her to leave the jail.

To explain what he means by a "meaningful" reduction of normative freedom, Yaffe provides the following definition:

> D suffers a meaningful reduction in normative freedom if and only if there is an act A such that (i) at time t1 it is permissible for D to A; (ii) at time t2 it is impermissible for D to A; (iii) after t2, D does not A; and (iv) the fact that D does not A after t2 is explained, at least in part, by the fact that it is impermissible for him to A after t2.[118]

We can easily fill in this picture for pre-trial detention. At t1, namely before D was sent to pre-trial detention, it is permissible for D to A, to be outside the jail. At t2, namely once D is detained, it is impermissible for D to be outside the jail. Moreover, at t2 while the D is detained, D is not outside the jail. Last, the fact that D is not outside the jail is explained by the fact that it is legally impermissible for D to leave the jail after D was sent to pre-trial detention. Thus, there is a meaningful reduction of normative freedom in cases of pre-trial detention.

The next part of the definition is that the meaningful reduction in normative freedom must be required. Yaffe uses the word "required" to indicate that no one else can do P for D. It rules out a proxy system for punishment, and it also rules out fines and other monetary penalties that can be paid on one's behalf by someone else. In the context of pre-trial detention, D cannot use a proxy by having another person take D's place in the jail while awaiting trial.[119] Pre-trial detention, therefore, meets the "required" component of the norm shift theory.

---

[117] Yaffe, "The Norm Shift Theory of Punishment," 489.
[118] Yaffe, 493.
[119] The "required" condition raises an interesting question about cases of pre-trial detention in which a person is only incarcerated because the person could not pay bail. In theory, one could argue that another person can post the bail for D, so the incarceration may not be "required" in the proper sense. I respond, however, that detention in such a case still

The second part of the punishment-by-nature definition is that P occurs thanks to government S's exercise of its punitive power. Yaffe provides the following explanation of punitive power:

> The punitive power is possessed only where there is a finding (perhaps accurate, perhaps not) that D committed C, where C is wrongful.[120]

Here, it is important to distinguish between punishment and *justified* punishment. The requirement for a "finding" in this context seems to be a requirement for something like a conviction. Though it does not need to be a proper conviction, convictions are still often understood as required for justified punishment. The finding may be a necessary condition for certain definitions of *justified* punishment, but it does not seem to be a necessary condition for all punishment, since there are instances of punishment that occur without a finding.

Consider the following examples of instances of what seems like punishment in the absence of a finding:

> **Robbers**: The state believes that three people committed a terrible robbery, which drained the state's treasury. State officials arrive at their homes, arrest them, and send them to prison for 30 years.
>
> **Protestors**: Due to dire economic conditions, thousands of protestors take to the streets with signs denouncing the government and its policies. At the protest, police round up a group of protestors and take them to a local jail. The government executes the protestors the next day.
>
> **Terrorists**: Following the worst terror attack in decades, the government finds people that it believes may have helped plan the attack. The government then tortures those people and incarcerates them for the rest of their lives.

In all three of these cases, it seems that punishment is occurring: in **Robbers**, people are imprisoned for 30 years; in **Protestors**, people are executed; and in **Terrorist**, people are tortured and incarcerated. Yet, in all three of these cases, there is no finding that those punished committed the crime. Thus, it seems that a finding may not be required for something to be punishment. If a finding is not required for a treatment to be punishment, then the lack of conviction in cases of pre-trial detention does not preclude it from being punishment. Since it meets all of the other necessary conditions, one can conclude that pre-trial detention is punishment.

---

satisfies the "required" condition, since, once bail is not posted, no one else can take D's place in jail. Thus, the mandate to post bail would not constitute punishment but subsequent detention would.

[120] Yaffe, "The Norm Shift Theory of Punishment," 503. Yaffe also requires that the punitive power have a target-hardening entitlement norm, which essentially means that the state can use force to enforce the norm (i.e., to incarcerate a person). He adds this component to differentiate punishment-by-nature from civil injunctions. Pre-trial detention meets that norm since the people are forcibly incarcerated pre-trial.

One may respond, however, that in all three of those cases, the state has a belief or a suspicion that the person committed the crime. Even if the state never formally announces that belief, perhaps this belief is sufficient to meet the finding requirement.

I have two responses: First, it seems that even a belief or a suspicion that a person committed the crime is not necessary for something to be punishment. Consider the following case, which is adapted from Charles Nesson's famous case:[121]

> **Prison Yard**: 25 people are exercising in a prison yard, when 20 of them suddenly join together in a planned attack on the prison guards. The remaining five people do not participate in the attack in any way. In response to this incident, the state adds five years to everyone's sentence, including the five who did not participate.

In this case, it seems that the five additional years of incarceration added to each sentence is punishment.[122] Yet, in five cases, the state does not have belief or a suspicion that the person participated in the attack. It still seems as though the state is punishing those five people, even though the state believes that those five people did not break the law or commit a crime. Thus, it seems that there are cases of punishment in which there is no belief or suspicion that the person committed a crime.

Notably, this type of case, in which the state knows that the person is innocent, aligns with cases in the literature that are referred to as punishment. For instance, a common objection to a consequentialist theory of punishment is that it may justify the following scenario: the state knowingly incarcerates an innocent person to avert an angry mob.[123] Hart specifically mentions the importance of designating those cases as punishments, even if they may not be justified punishments.

Second, if the bar for a finding is merely a belief or a suspicion that a person committed a crime, then pre-trial detention certainly meets that requirement. A person is detained pre-trial after the person has been charged with a crime. A charge is the state's pronouncement that it believes that the person committed a crime. Thus, it seems that in instances of pre-trial detention, the state does possess a belief or a suspicion that the person committed a crime.

Moreover, even if one adopted a more limited understanding of a finding, pre-trial detention would still fit the definition. A charge is much more than an announcement of suspicion. Rather, the state must provide sufficient evidence to a judge or a grand jury in order to charge someone with a

---

[121] Nesson, "Reasonable Doubt and Permissive Inferences."
[122] This case could be described as an instance of collective punishment.
[123] See, e.g., Lippke, "Punishing the Guilty, Not Punishing the Innocent."

crime. A charge has many of the same features of conviction—the primary difference is the standard of proof required. But both follow the presentation of evidence to an arbiter, either a judge or a jury. In the US, for instance, there are cases of both charges and convictions that are determined by judges, and there are cases of both that are determined by juries. Importantly, the amount of evidence and level of review required for charges in jurisdictions like the US and the UK may even be higher than those required for convictions in other jurisdictions. In certain jurisdictions, post-"trial" sentences are meted out with fairly little evidence and review. For instance, people were arrested in late 2022 for protesting the regime in Iran following the death of Masa Amini.[124] Four of them were given 15 minutes in court to defend themselves before they were sentenced to death and executed. In these cases, the death penalty is certainly punishment. Yet, it was imposed following a sham trial. Far less evidence was presented at their trials than is typically presented to a judge or a grand jury to secure a charge.

One may reply, however, that there is an important distinction between a charge and a conviction, making treatment following a charge non-punishment and treatment following a conviction punishment. One could argue that the state has a mere belief when it charges a person with a crime, but in the context of conviction, the state has knowledge (or at least believes it has knowledge).

I reply, however, that this distinction does not actually differentiate between charges and convictions. Consider a modified version of the **Prison Yard** case:

> **Prison Yard***: 25 prisoners are exercising in a prison yard, when 24 of them suddenly join together in a planned attack on the prison guards. The remaining prisoner tries to stop the attack. There is no available evidence distinguishing the innocent prisoner from the rest. Local prosecutors randomly select one of the prisoners. That person is tried, convicted, and sentenced to 10 additional years of incarceration.

In this case, the conviction is based on statistical evidence.[125] The jury believes that there is a 96% chance that the person at trial committed the crime. The jury knows that they cannot be certain that the person participated in the attack. But, on the basis of the statistical evidence alone, the jury decides to convict anyways. While it may be an improper conviction, it is nonetheless a conviction. The 10 additional years of incarceration are certainly punishment. Thus, it seems that punishment can follow convictions on the basis of a mere belief. If so, then the concern that charges do not amount to knowledge (or belief that one possesses knowledge) does not stand.

---

[124] "Iran Protests."

[125] I return to the issue of conviction on the basis of statistical evidence later in this thesis.

To conclude, we have two options regarding the finding requirement: either we remove it from the definition or we keep it in. If we remove it, pre-trial detention meets the conditions of the definition. If we keep it in, pre-trial detention still meets the conditions of the definition. Thus, pre-trial detention seems to fit the punishment-by-nature definition provided by Yaffe's Norm Shift Theory of Punishment.

To review, I demonstrated that pre-trial detention fits Hart's canonical definition of punishment and Yaffe's Norm Shift Theory. I therefore conclude that the current practice of pre-trial detention is punishment.

§5. Permissibility

Now that we have seen that pre-trial detention is punishment, we can clearly see that it is impermissible punishment. Quite straightforwardly, the practice is unjust, since it systematically inflicts punishment before people have been tried. It is also unconstitutional, which is why the US Supreme Court goes to great lengths to call it "not punishment."

There is a further implication to clarifying that pre-trial detention is punishment. It helps us address the following question: Aren't there cases of pre-trial detention that are permissible? Though it is often not directly articulated, a version of this argument animates much of the literature on pre-trial detention:

(1) Punishment before trial is not permissible.

(2) Pre-trial detention is permissible.

Conclusion: Pre-trial detention is not punishment.

From this argument, we can see why scholars would want to shy away from calling pre-trial detention punishment.[126] In support of (2), consider a serial killer who has a slew of victims, or a complex case in which the evidence against a mob boss is mysteriously disappearing. In these cases, it seems that it may be permissible for the state to incarcerate the serial killer and the mob boss before trial. Since punishment without a trial is not permissible, many conclude that pre-trial detention must not be punishment.

Yet, once we see that the current practice of pre-trial detention is in fact punishment, we see that there may be two different types of permissibility in this argument, which leads to confusion. The term "permissible" in (1) denotes permissible *as* punishment, while in (2) the same word denotes

[126] See, e.g., Lippke, "Preventive Pre-Trial Detention without Punishment," 114; Mayson, "Dangerous Defendants," 551.

permissible as something else. For instance, the detention of a person before trial could be permissible as medical quarantine, involuntary commitment, forced conscription, jury sequestration, etc. To demonstrate this, consider the following case:

> After D is charged with a crime, D returns home to await trial. Soon after, she is infected by a highly contagious virus. She is then taken to a quarantine facility and held there against her will.

In such a scenario, D is detained by the state before her trial and thus experiences pre-trial detention. Yet, this detention is not punishment. In terms of Hart's definition, it is not "for an offence" but rather is a result of her contagious virus. Thus, there could be cases of incarceration before trial that are permissible, but not permissible as punishment.

The question of whether it would be permissible to incarcerate the serial killer or the mob boss would follow the question of whether it would be permissible to incarcerate a person who had no connection to a crime on the same basis. D's quarantine is permissible only insofar as it is permissible for the state to quarantine anyone with a contagious disease.[127] The quarantine is completely unrelated to the fact that D has been charged with a crime. The state would quarantine anyone with that contagious virus, which would primarily be people who have never been charged with a crime. Indeed, D's quarantine facility would likely be primarily filled with people who are not awaiting trial.

Similarly, in order for the detention of the serial killer and the mob boss to not be punishment, the same justification would need to be able to cover the detention of a person who is neither suspected of nor charged with a crime. Detaining the serial killer would be a form of preventive detention, which is when the state preemptively detains a person on the grounds that the person poses a future threat that needs to be averted. Preventive detention is meant to refer to cases in which a person has not been charged with a crime. Thus, if it is permissible to detain the serial killer, then it must also be permissible to subject anyone to preventive detention who poses a similar threat, even if the person has not already done anything illegal or impermissible.

Notably, many scholars think that preventive detention is impermissible. Doug Husak, for instance, argues that preventive detention is *only* permissible when imposed as a standard legal punishment.[128] He argues that posing a threat should count as committing a crime and that preventive

---

[127] For an argument regarding compensation for quarantine detention, see Holm, "Should Persons Detained During Public Health Crises Receive Compensation?"

[128] Husak, "Preventive Detention as Punishment?: Some Possible Obstacles."

detention should be the punishment for that crime. Indeed, he argues that all procedural safeguards in place for post-trial punishment (e.g., right to fair trial, presumption of innocence, right to remain silent) should apply to cases of preventive detention. Husak's view is that preventive detention without this framework and without these safeguards is impermissible. If Husak is correct, then preventatively detaining the serial killer before trial would be impermissible.

Similarly, it would be permissible to incarcerate the mob boss only if it is permissible to incarcerate anyone who might destroy evidence needed for a criminal investigation. This would mean that the state could permissibly incarcerate any person that the state suspects would destroy or alter evidence. Family and friends of a criminal defendant, as well as family and friends of a victim, could be permissibly detained if the state reasonably suspects that they may tamper with evidence, even if they have not (yet) done anything illegal. Permitting the state to incarcerate people merely on the basis of suspicion seems fairly dubious. Regardless, the permissibility of detaining the mob boss will rise and fall with the permissibility of detaining anyone, including his friends and family, on those same grounds.

But, since the current practice of pre-trial detention is punishment, there is actually no way to permissibly incarcerate people in a jail while awaiting trial. To demonstrate why this is the case, let us consider the following case:

> E commits a crime but is found to be too mentally incompetent to be held responsible for her actions. Following trial, E is sent to a prison instead of to a mental health facility.

When a person is severely mentally ill and cannot be held responsible for her actions, punishment is never permissible. It may be the case that the state could permissibly detain E at a mental health facility against her will. But it is still impermissible for the state to punish E. The state, therefore, wrongs E by sending her to a prison, even if the state could have permissibly detained E at a mental health facility. This case is analogous to those of the serial killer and the mob boss. When a person is awaiting trial, incarceration is never permissible as punishment. It may be the case that the state could permissibly detain the serial killer or mob boss in the same way that the state would detain a person who was not accused of a crime. But it is impermissible for the state to punish them. The state, therefore, wrongs the serial killer and mob boss by sending them to jail, even if the state could have permissibly detained them. Thus, in a jail, everyone, even those whom the state may have otherwise permissibly incarcerated, is impermissibly punished.

One could respond that perhaps everyone in jail is similar to the serial killer or mob boss, meaning that the state has a reason to permissibly detain them. Perhaps, then, jail is just the place that detention for those reasons takes place.

I have two responses. First, as we have seen, the current practice of pre-trial detention is closely tied to being charged with a crime. Immediately after a person is formally charged with a crime, a judge decides whether or not the person should be sent to pre-trial detention. Pre-trial detention is a step in a larger criminal process that spans from arrest through conviction. The detention is "for the offence" listed in the charge, and the time spent in detention is later credited in a 1:1 ratio towards a sentence for that charge specifically. Thus, the current practice of pre-trial detention is specifically predicated on the person being charged with a crime.

Second, if pre-trial detention in jails was, in fact, not related to being charged with a crime, then one would expect to find at least some people who are not charged with a crime incarcerated in jails. In D's quarantine facility, for instance, there may be a mix of people who have and who have not been charged with crimes. Similarly, if E had been properly sent to a mental health facility, there also would have been a mix of people who had and had not committed crimes. Yet, everyone in jail has been charged with a crime. Indeed, it is no surprise that we do not find people in jail who have not been charged with a crime; incarcerating people in jail who have done nothing illegal would be highly problematic.

To review, this section addressed the unsettled question regarding the permissibility of certain cases of pre-trial detention. Though it may be permissible for a person to be charged with a crime and detained for other reasons (e.g., quarantine, mental health), the current practice of pre-trial detention is impermissible in all cases. Even if the state would have otherwise had a reason to permissibly incarcerate someone, by sending that person to jail, the state wrongly punishes that person.

§6. Credit & Compensation Model

I began this paper by arguing that the practice of crediting time served indicates that pre-trial detention is punishment. Now that we have clarified that pre-trial detention is unjustified punishment, where do we stand on the practice of crediting time served?

In the literature, Kolber and Donelson support the practice of crediting time served.[129] Ferzan, however, argues in favor of abolishing the practice of crediting time served. Instead, she proposes that

[129] Kolber, "Against Proportional Punishment"; Donelson, "Natural Punishment."

all detainees—both those who will later be convicted and those who will later be acquitted—should be compensated in the same way and for the same reason that we compensate jurors.[130] In this section, I propose a hybrid view that includes both credit and compensation. Given my arguments above, my explanation as to *why* the state should credit and compensate for time served is distinct from the explanations found in the literature.

6.1 Convicted Defendants

I argue that people who are subject to pre-trial punishment and then convicted should receive *both* a credit for the time served *and* compensation. Time served in jail awaiting trial should be credited for the reason discussed throughout this paper: it is part of the overall punishment. If the punishment begins when a person is charged with a crime and then detained, then the length of the sentence should be counted from there.

People convicted of a crime who were punished pre-trial should also be compensated for the impermissible punishment that they experienced. If legal guilt is a necessary condition for permissible punishment, then pre-trial punishment is definitionally never permissible punishment. Thus, the state is wronging the person by impermissibly punishing them. If compensation is owed when one party wrongs another, then the state should compensate the pre-trial detainees it is wronging. Even if the person turns out to be factually guilty, the state did not have the moral license to punish the person before conviction. This argument also applies to people who, like the serial killer, mob boss, or contagious patient, the state could have otherwise permissibly incarcerated. Just as the state wrongs E by sending her to prison instead of a mental health facility, so too the state wrongs these individuals by sending them to jail.

One may argue, however, that the credit towards the post-sentence incarceration is sufficient compensation. With the credit in place, the person is no worse off than she would have been otherwise, since she ultimately spends the same number of years incarcerated. Thus, perhaps the credit for time served sufficiently compensates the detainee for the impermissible punishment.

I have two responses: (1) people are worse off when part of their punishment occurs pre-trial, and (2) even if the person's sentence is ultimately the same length, something more is owed, given that the part of the punishment which occurred pre-trial was impermissible.

[130] Ferzan, "The Trouble with Time-Served."

There are a number of reasons why people are worse off when serving part of their sentence pre-trial as opposed to serving their full sentence post-trial. First, the conditions in jails are generally far worse than those in prisons.[131] Second, pre-trial punishment can interfere with the ability to prepare properly for trial. For instance, it may be more difficult to meet with attorneys and construct a defense for the case.[132] Studies have shown that defendants are more likely to get convicted, all else equal, if they await trial in jail instead of at home.[133] Third, the anxiety and uncertainty of awaiting trial can make incarceration during that period even more painful and difficult.[134]

Second, even if the credit means that the person is just as well off as she would have been without any pre-trial punishment, it is important to recognize that the state is wronging the person by imposing impermissible punishment. This remains true even if the state has the correct belief that the person will later be convicted. The need for additional compensation is demonstrated through this unrelated, but structurally similar, case:

> F needs $100 and sees G's wallet unattended. F believes that there is a good chance that G may later owe F money as a result of the poker game later that day. F steals $100 from G. Later on, G loses the poker game and owes F $150.

Certainly, G should only have to give F $50 at most. But it seems that F would still owe additional compensation or an apology, even though F and G end up with the correct amount of money. Similarly, even if a person is incarcerated for the correct number of years overall, it seems that the state would still owe additional compensation or an apology. Simply restoring the person to her state before the wrongdoing occurred is insufficient. Joel Feinberg, for instance, argues that "compensation for harm wrongly inflicted by others…functions not only to repair the damages but also to 'restore the moral equilibrium' as would an apology or expression of remorse."[135] Thus, only crediting time served is insufficient. Instead, monetary compensation in addition to the credit for time served, or a

---

131 Vansickle and Villa, "Who Begs to Go to Prison? California Jail Inmates."

132 See, e.g., Dobbie, Goldin, and Yang, "The Effects of Pretrial Detention on Conviction, Future Crime, and Employment"; Heaton, Stevenson, and Mayson, "The Downstream Consequences of Misdemeanor Pretrial Detention"; Stevenson, "Distortion of Justice."

133 See, e.g., Leslie and Pope, "The Unintended Impact of Pretrial Detention on Case Outcomes"; Dobbie, Goldin, and Yang, "The Effects of Pretrial Detention on Conviction, Future Crime, and Employment"; Heaton, Stevenson, and Mayson, "The Downstream Consequences of Misdemeanor Pretrial Detention"; Stevenson, "Distortion of Justice."

134 In the US, the suicide rate in jails is three times the suicide rate in prisons, which is already higher than the suicide rate for the general population. See Cain and Ellison, "Identifying Individuals at Risk of Suicide and Self-Harm in Jail." Studies have demonstrated that the uncertainty of awaiting trial can significant increases anxiety levels. See, e.g., Oleski, "The Effect of Indefinite Pretrial Incarceration on the Anxiety Level of an Urban Jail Population."

135 Feinberg, *Doing & Deserving*, 74.

further reduced sentence (e.g., two years of pre-trial punishment reduces post-trial punishment by four years), would be necessary in order to restore the moral equilibrium.

Moreover, the case for compensation in addition to crediting time served is similar to instances in which a person is harmed while serving a sentence. The harm is separate from the imposed sentence. In many jurisdictions, when a person is harmed while serving a prison sentence, the person can sue the state for compensation.[136] A person may be eligible for compensation from the state, even if she is not directly harmed by the state.[137] Suppose that J and K are both incarcerated, and J attacks K. K can sue the state for failing to protect her, even though she was harmed by J. Since K is in the state's custody, the state wrongs K by failing to protect her. In such cases, the overall length of incarcerated is not changed, but the state may owe compensation.[138] Similarly, when a person is punished pre-trial, the person still serves her overall sentence, spanning from pre-trial incarceration through post-trial incarceration. Yet, an additional wrong is occurring, and the state thus owes compensation for that wrong.

6.2 Acquitted Defendants

What should happen to people who are detained pre-trial and then released? People may be released for different reasons. In Browder's case, for instance, he was released because the charges against him were dropped. In other cases, people may be tried and then acquitted. In all of these scenarios, there is no post-trial punishment that can be reduced on account of the pre-trial punishment.

I argue that compensation is owed to those who are punished pre-trial and then released. In addition to their legal innocence, many of them may be factually innocent as well. Thus, the punishment they experience is akin to post-trial punishment of an innocent person. For the same reasons that the state should compensate innocent people who are wrongly punished post-trial, the state should compensate people who are wrongly punished pre-trial and then released.[139]

---

[136] In the US, the two Supreme Court established a right to compensation in such a case. See Bivens v. Six Unknown Fed. Narcotics Agents, 403 U.S. 388 (1971); Carlson v. Green, 446 U.S. 14 (1980). For a descriptions of the resulting jurisprudence in the US, see, e.g., Champion, "Some Recent Trends in Civil Litigation by Federal and State Prison Inmates"; Frink, "Defining a Standard for Punitive Damages in Prisoners' Rights Litigation"; Mahoney, "Punitive Damage and the Use of Modern Common Law in Construing Section 1983: Smith v. Wade."

[137] For instance, people have sued for compensation and punitive damages when family members died from COVID-19 while incarcerated. See, e.g., Clayton, "Family of Man Who Died from Covid While Incarcerated Sues California Prison System."

[138] See, e.g., Edwards v United Kingdom (2002) 35 EHRR 487 (ECHR). In that case, the European Court of Human Rights found that the UK owed compensation to the family of an incarcerated person who was killed by a cellmate.

[139] I return to the question of state-owed compensation for wrongful punishment later in this thesis.

One may be concerned, however, that the practice of compensating both convicted and acquitted defendants, while only crediting convicted defendants, leads to an unfair system. If convicted detainees are both compensated and credited, then they will be better off than acquitted detainees who are only compensated.

I respond, however, that the degree of compensation can and should differ, depending on the harm that the compensation is meant to repair. If an acquitted detainee is receiving compensation for pre-trial punishment, the amount would reflect the length of the pre-trial punishment, in addition to whatever is further needed to restore the moral equilibrium. The compensation for a person who is punished pre-trial and then convicted, however, would only reflect what is needed to restore the moral equilibrium. Thus, if an acquitted defendant and a convicted defendant are punished pre-trial for the same amount of time, the acquitted defendant would be owed much more compensation than the convicted defendant would be owed.

6.3 Implications of the Compensation & Credit Model

Thus far, I have argued that pre-trial detention is pre-trial punishment. I then proposed a hybrid compensation and credit model. I will now briefly outline the implications of this model and clarify how it is distinct from the solutions previously suggested by scholars.

First, my view provides a cohesive and consistent explanation of what is occurring. It does not require a temporal switch—it is punishment both while it occurs and when it comes to sentencing. It also describes the treatment of pre-trial detainees who are acquitted and of those who are convicted in the same manner. The compensation and credit model ensures that both groups are addressed, unlike the current practice of only crediting convicted defendants. Thus, my view provides an explanation that remains consistent over time and consistent for all defendants.

Second, my view provides a path forward, given the current conditions of pre-trial detention. Ferzan's model, for instance, requires a dramatic re-envisioning of the practice of pre-trial detention. Specifically, she notes that abolishing credit for time served and replacing it with monetary compensation only makes sense if jail conditions are dramatically different than they are today.[140] While I wholeheartedly agree that the conditions of US jails must change, the plausibility of my view does not rely on jails looking dramatically different than they do today. Thus, my view works on its own, right now, even if jail conditions do not change.

[140] Ferzan, "The Trouble with Time-Served," 75.

Third, my view provides a motivation for limiting the practice of pre-trial detention. Currently, the practice of detaining people pre-trial and then crediting time served often makes imposing a sentence overall less costly for the state. In the US, for instance, jails tend to be cheaper to run (in part because of their poor conditions), so states often save money by having people serve part of their sentences pre-trial in a cheaper facility.[141] Abolishing credit for time served and replacing it fully with compensation, per Ferzan's suggestion, would be much more costly for the state and thus would serve as a motivation to reduce pre-trial detention.[142] The issue with the proposal, however, is that people are then incarcerated for far longer than they should be. My view both maintains credit for time served and sets up a financial disincentive for punishing people pre-trial.

Conclusion

In this paper, I argued that pre-trial detention is punishment. After laying out the two puzzles regarding crediting time served, I demonstrated how credit for time served provides a presumptive argument for understanding pre-trial detention as punishment, which in turn resolves both puzzles. I then argued that pre-trial detention meets key definitions of punishment. Finally, I proposed a hybrid compensation and credit model for those punished before trial. Though the compensation and credit model would benefit from further development regarding its application on a case-by-case basis, my view better reflects the realities of our current practice and sets out a path forward for a more just practice.

Bibliography

Alschuler, Albert W. "Preventive Pretrial Detention and the Failure of Interest-Balancing Approaches to Due Process." *Michigan Law Review* 85, no. 3 (1986): 510–69. https://doi.org/10.2307/1288756.

Armstrong, K.J. "The Retributivist Hits Back." *Mind* LXX, no. 280 (October 1961): 471–90. https://doi.org/10.1093/mind/LXX.280.471.

Ashworth, Andrew. "Four Threats to the Presumption of Innocence." *The International Journal of Evidence & Proof* 10, no. 4 (July 2006): 241–79. https://doi.org/10.1350/ijep.10.4.241.

Baradaran, Shima. "Restoring the Presumption of Innocence." *Ohio State Law Journal* 72, no. 4 (2011): 723–76.

———. "The Presumption of Punishment." *Criminal Law and Philosophy* 8, no. 2 (2014): 391–406. https://doi.org/10.1007/s11572-013-9236-7.

[141] Since jails are often run by local governments, the state reimburses the jail per person incarcerated. This reimbursement rate, however, is often lower than the price per person to incarcerated someone in a prison. Henrichson and Delaney, "The Price of Prisons," 70.

[142] Ferzan, "The Trouble with Time-Served."

BBC News. "Iran Protests: 15 Minutes to Defend Yourself against the Death Penalty." January 18, 2023, sec. Middle East. https://www.bbc.com/news/world-middle-east-64302726.

Bedau, Hugo Adam. "Retribution and the Theory of Punishment." *The Journal of Philosophy* 75, no. 11 (1978): 601–20.

Benn, S. I. "An Approach to the Problems of Punishment." *Philosophy* 33, no. 127 (1958): 325–41.

Berman, Mitchell N. "Punishment and Justification." *Ethics* 118, no. 2 (January 1, 2008): 258–90. https://doi.org/10.1086/527424.

Bisogno, Enrico, Angela Me, Maurice Dunaiski, and Theodore Leggett. "Data Matters." United Nations Office on Drugs and Crime, 2021.

Bivens v. Six Unknown Fed. Narcotics Agents, 403 U.S. 388 (1971) (June 21, 1971).

Blagojević & Jokić (IT-02-60), No. IT-02-06 (International Criminal Tribunal for the former Yugoslavia May 9, 2007).

Boonin, David. *The Problem of Punishment*. Cambridge: Cambridge University Press, 2008. https://doi.org/10.1017/CBO9780511819254.

Brunt, Alexa Van, and Locke E Bowman. "Toward a Just Model of Pretrial Release: A History of Bail Reform and a Prescription for What's Next." *Journal of Criminal Law and Criminology* 108, no. 4 (2018): 701–74.

Burgh, Richard W. "Do the Guilty Deserve Punishment?" *The Journal of Philosophy* 79, no. 4 (1982): 193–210. https://doi.org/10.2307/2026220.

Burns, Ronald. "Assessing Jail Coverage in Introductory Criminal Justice Textbooks." *Journal of Criminal Justice Education* 13, no. 1 (March 1, 2002): 87–100. https://doi.org/10.1080/10511250200085341.

Cain, Calli M, and Jared M. Ellison. "Identifying Individuals at Risk of Suicide and Self-Harm in Jail." *Corrections* 0, no. 0 (January 22, 2022): 1–16. https://doi.org/10.1080/23774657.2022.2031350.

Carlson v. Green, 446 U.S. 14 (1980) (April 22, 1980).

Champion, Dean J. "Some Recent Trends in Civil Litigation by Federal and State Prison Inmates." *Federal Probation* 52, no. 3 (1988): 43–47.

Clayton, Abené. "Family of Man Who Died from Covid While Incarcerated Sues California Prison System." *The Guardian*, March 18, 2021, sec. US news. https://www.theguardian.com/us-news/2021/mar/18/california-coronavirus-prison-system-lawsuit.

Criminal Justice Act 2003 s.240ZA Credit for Time Spent on Remand (n.d.). Accessed November 20, 2022.

Defence-Barrister.co.uk. "Credit for Time Spent on Remand." Accessed November 1, 2020. https://www.defence-barrister.co.uk/new-page.

Dobbie, Will, Jacob Goldin, and Crystal S. Yang. "The Effects of Pretrial Detention on Conviction, Future Crime, and Employment: Evidence from Randomly Assigned Judges." *American Economic Review* 108, no. 2 (2018): 201–40. https://doi.org/10.1257/aer.20161503.

Donelson, Raff. "Cruel and Unusual What? Toward a Unified Definition of Punishment." *Washington University Jurisprudence Review* 9, no. 1 (2016): 1–41.

———. "Natural Punishment." *North Carolina Law Review* 100, no. 2 (January 2022): 557–600.

Duff, R. A. "Pre-Trial Detention and the Presumption of Innocence." In *Prevention and the Limits of the Criminal Law*, edited by Andrew Ashworth, Lucia Zedner, and Patrick Tomlin, 115–32. Oxford University Press, 2013. https://doi.org/10.1093/acprof:oso/9780199656769.003.0007.

———. *Punishment, Communication, and Community*. Oxford University Press, 2000.

Edwards v United Kingdom (2002) 35 EHRR 487 (ECHR) (March 14, 2002).

Feinberg, Joel. *Doing & Deserving: Essays in the Theory of Responsibility*. Princeton University Press, 1970.

———. "The Expressive Function of Punishment." *The Monist* 49, no. 3 (1965): 397–423.

Fellner, Jamie. "The Price of Freedom: Bail and Pretrial Detention of Low Income Nonfelony Defendants in New York City." New York, N.Y.: Human Rights Watch, 2010.
Ferzan, Kimberly Kessler. "Beyond Crime and Commitment: Justifying Liberty Deprivations of the Dangerous and Responsible." *Minnesota Law Review* 96 (2011): 141–93.
———. "Preventive Justice and the Presumption of Innocence." *Criminal Law and Philosophy* 8, no. 2 (June 1, 2014): 505–25. https://doi.org/10.1007/s11572-013-9275-0.
———. "The Trouble with Time-Served." *Brigham Young University Law Review*, forthcoming. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4041973.
Flew, Antony. "The Justification of Punishment." *Philosophy* 29, no. 111 (1954): 291–307. https://doi.org/10.1017/s0031819100067152.
"For, Prep., Conj., n., and Adv." In *Oxford English Dictionary*. Oxford University Press, December 2022. https://www.oed.com/view/Entry/72761.
Frink, Karen A. "Defining a Standard for Punitive Damages in Prisoners' Rights Litigation: Smith v. Wade Case." *Syracuse Law Review* 35, no. 3 (1984): 1037–66.
Gonnerman, Jennifer. "Kalief Browder, 1993–2015." The New Yorker. Accessed February 17, 2020. https://www.newyorker.com/news/news-desk/kalief-browder-1993-2015.
Hart, H. L. A. *Punishment and Responsibility: Essays in the Philosophy of Law*. Oxford: Clarendon P., 1968.
Heard, Catherine. "Pre-Trial Detention and Its Overuse," November 11, 2019. http://blogs.bbk.ac.uk/research/2019/11/11/pre-trial-detention-and-its-overuse/.
Heaton, Paul, Megan Stevenson, and Sandra Mayson. "The Downstream Consequences of Misdemeanor Pretrial Detention." *Stanford Law Review* 69, no. 3 (2017): 711–94. https://doi.org/10.2139/ssrn.2809840.
Henrichson, Christian, and Ruth Delaney. "The Price of Prisons: What Incarceration Costs Taxpayers." *Federal Sentencing Reporter* 25, no. 1 (2013 2012): 68–80.
Holm, Søren. "Should Persons Detained During Public Health Crises Receive Compensation?" *Journal of Bioethical Inquiry* 6, no. 2 (June 1, 2009): 197–205. https://doi.org/10.1007/s11673-009-9160-7.
Husak, Douglas. "A Framework for Punishment: What Is the Insight of Hart's 'Prolegomenon'?" In *Hart on Responsibility*, edited by Christopher Pulman, 91–108. Philosophers in Depth. London: Palgrave Macmillan UK, 2014. https://doi.org/10.1057/9781137374431_5.
———. "A Framework for Punishment: What Is the Insight of Hart's 'Prolegomenon'?" In *Hart on Responsibility*, edited by Christopher Pulman, 91–108. Philosophers in Depth. London: Palgrave Macmillan UK, 2014. https://doi.org/10.1057/9781137374431_5.
———. "Already Punished Enough." *Philosophical Topics* 18, no. 1 (1990): 79–99.
———. "Lifting the Cloak: Preventive Detention as Punishment." *San Diego Law Review* 48 (2011): 1173–1204.
———. "Preventive Detention as Punishment?: Some Possible Obstacles." In *Prevention and the Limits of the Criminal Law*, edited by Andrew Ashworth, Lucia Zedner, and Patrick Tomlin. Oxford University Press, 2013. https://oxford.universitypressscholarship.com/view/10.1093/acprof:oso/9780199656769.001.0001/acprof-9780199656769-chapter-10.
Klein, Douglas J. "The Pretrial Detention 'Crisis': The Causes and the Cure." *Journal of Urban and Contemporary Law* 52 (January 1997): 27.
Kolber, Adam J. "Against Proportional Punishment." *Vanderbilt Law Review* 66, no. 4 (2013): 1141–79.
———. "The Comparative Nature of Punishment." *Boston University Law Review* 89, no. 5 (2009): 1565–1608.

———. "The Subjectivist Critique of Proportionality." In *The Palgrave Handbook of Applied Ethics and the Criminal Law*, edited by Larry Alexander and Kimberly Kessler Ferzan, 571–95. Cham: Springer International Publishing, 2019. https://doi.org/10.1007/978-3-030-22811-8_24.

Lacey, Nicola. *State Punishment: Political Principles and Community Values*. Psychology Press, 1994.

Leslie, Emily, and Nolan G. Pope. "The Unintended Impact of Pretrial Detention on Case Outcomes: Evidence from New York City Arraignments." *The Journal of Law and Economics* 60, no. 3 (August 2017): 529–57. https://doi.org/10.1086/695285.

Lippke, Richard L. "Preventive Pre-Trial Detention without Punishment." *Res Publica* 20, no. 2 (May 2014): 111–27. https://doi.org/10.1007/s11158-013-9234-6.

———. "Punishing the Guilty, Not Punishing the Innocent." *Journal of Moral Philosophy* 7, no. 4 (January 1, 2010): 462–88. https://doi.org/10.1163/174552410X511446.

Mahoney, Edward F. "Punitive Damage and the Use of Modern Common Law in Construing Section 1983: Smith v. Wade." *Boston College Law Review* 25, no. 5 (1984): 1001–27.

Makar, Zina. "Unnecessary Incarceration." *Oregon Law Review* 98, no. 2 (2020): 607–70.

Manns, Jeffrey. "Liberty Takings: A Framework for Compensating Pretrial Detainees." *Cardozo Law Review* 26 (2000): 1947–2024.

Mayson, Sandra. "Dangerous Defendants." *Yale Law Journal* 127, no. 3 (2017): 490–569.

McPherson, Thomas. "Punishment: Definition and Justification." *Analysis* 28, no. 1 (1967): 21–27. https://doi.org/10.2307/3327609.

Miller, Marc, and Martin Guggenheim. "Pretrial Detention and Punishment." *University of Minnesota Law Review* 75 (1990): 93.

Myers, Nicole Marie. "Eroding the Presumption of Innocence: Pre-Trial Detention and the Use of Conditional Release on Bail." *The British Journal of Criminology* 57, no. 3 (2017): 664–83. https://doi.org/10.1093/bjc/azw002.

Nesson, Charles R. "Reasonable Doubt and Permissive Inferences: The Value of Complexity." *Harvard Law Review* 92, no. 6 (1979): 1187–1225. https://doi.org/10.2307/1340444.

Oleski, Merritt S. "The Effect of Indefinite Pretrial Incarceration on the Anxiety Level of an Urban Jail Population." *Journal of Clinical Psychology* 33, no. 4 (1977): 1006–8.

Pulman, Christopher, ed. *Hart on Responsibility*. London: Palgrave Macmillan UK, 2014. https://doi.org/10.1057/9781137374431.

Raz, Joseph. "H. L. A. Hart (1907–1992)." *Utilitas* 5, no. 2 (November 1993): 145–56. https://doi.org/10.1017/S0953820800005732.

Ristroph, Alice. "How (Not) to Think Like a Punisher." *Florida Law Review* 61 (2009): 727.

Samek, Robert A. "Punishment: A Postscript to Two Prolegomena." *Philosophy* 41, no. 157 (July 1966): 216–32. https://doi.org/10.1017/S0031819100058691.

Sifferd, Katrina L., and Tyler K. Fagan. "Pretrial Detention and Moral Agency." In *The Palgrave Handbook of Philosophy and Public Policy*, edited by David Boonin, 11–23. Cham: Springer International Publishing, 2018. https://doi.org/10.1007/978-3-319-93907-0_2.

Slifer, Stephanie. "How the Wrongfully Convicted Are Compensated for Years Lost." CBS News, March 27, 2014. https://www.cbsnews.com/news/how-the-wrongfully-convicted-are-compensated/.

Smilansky, Saul. "The Time to Punish." *Analysis* 54, no. 1 (1994): 50–53. https://doi.org/10.2307/3328104.

Spjut, R. J. "Criminal Law, Punishment, and Penalties." *Oxford Journal of Legal Studies* 5, no. 1 (1985): 33–46.

Stevenson, Megan T. "Distortion of Justice: How the Inability to Pay Bail Affects Case Outcomes." *The Journal of Law, Economics, and Organization* 34, no. 4 (November 1, 2018): 511–42. https://doi.org/10.1093/jleo/ewy019.

Stevenson, Megan T, and Sandra G Mayson. "Pretrial Detention and the Value of Liberty." *Virginia Law Review* 108 (2022): 74.

Sverdlik, Steven. "Punishment." *Law and Philosophy* 7, no. 2 (1989 1988): 179–202.

United States v. Salerno 481 U.S. 739 (1987) (May 26, 1987).

Vansickle, Abbie, and Manuel Villa. "Who Begs to Go to Prison? California Jail Inmates." Marshall Project. Accessed November 20, 2022. https://www.themarshallproject.org/2019/04/23/who-begs-to-go-to-prison-california-jail-inmates.

Wagner, Peter, and Wendy Sawyer. "Mass Incarceration: The Whole Pie 2023." Prison Policy Initiative, March 14, 2023. https://www.prisonpolicy.org/reports/pie2023.html.

Wiseman, Samuel R. "Pretrial Detention and the Right to Be Monitored." *The Yale Law Journal* 123, no. 5 (2014): 1344–1404.

Yaffe, Gideon. "The Norm Shift Theory of Punishment." *Ethics* 132, no. 2 (January 2022): 478–507. https://doi.org/10.1086/716874.

Criminal Confessions and Content-Sensitive Testimonial Injustice

Criminal confessions are often considered the "gold standard" of legal evidence. Yet, many of the people who are exonerated through DNA evidence originally falsely confessed to the crime. Juries frequently place excessive weight on the confessions of criminal defendants, which is epistemically puzzling. This paper draws on cases of excessive reliance on false confessions to make two interconnected contributions to the literatures on epistemic injustice and criminal confessions, respectively. First, I critique and refine Miranda Fricker's theory of testimonial injustice. I argue for an understanding of testimonial injustice that can clarify (1) the types of prejudices that can lead to testimonial injustice, (2) the debate regarding whether credibility excess cases can constitute instances of testimonial injustice, and (3) a previously unrecognized aspect of testimonial injustice, which I term content sensitivity. Second, with a clarified understanding of testimonial injustice, I argue that criminal defendants—both those who are factually innocent and those who are factually guilty—may experience testimonial injustice at trial. I conclude by considering some of the non-epistemic harms that result from these instances of testimonial injustice during criminal trials.

Introduction

After a four-day interrogation, Juan Rivera confessed to raping and killing an 11-year-old child when he nodded his head and signed a statement of confession. The DNA evidence at the scene, however, did not match Rivera's DNA. To account for both the confession and the unmatched DNA, the prosecution constructed the following theory of the case: the 11-year-old victim had consensual sex with an unidentified man. Rivera then arrived at the house, raped her, failed to ejaculate, and killed her—all while leaving no DNA traces. The jury was presented with this bizarre theory of the case, Rivera's confession, and DNA evidence that suggested that Rivera had not committed the crime. Yet, the jury convicted Rivera. When he was later re-tried two other times due to procedural errors, the two other juries also convicted on the basis of this evidence.[143]

[143] Loevy, "Juan Rivera's Coerced False Confession." All future references to this case are from here.

In this case, Rivera falsely confessed to a crime that he did not commit.[144] The *excessive* weight the juries placed on Rivera's confession—contrary to the DNA evidence—is not unusual. In fact, approximately 30% of those who have been exonerated through DNA evidence in the US originally falsely confessed to the crime.[145] Though in some of those cases the jury may have weighed the confession properly, it is likely that excessive weight was placed on defendants' confessions in many of those cases.[146] Indeed, studies have shown that jurors generally weigh criminal confessions—both true and false ones—far more heavily than other types of evidence.[147] Due to the frequent over-reliance on false confessions, some legal scholars have recently proposed that all confessions be excluded in criminal trials.[148]

The over-reliance on Rivera's confession, and on criminal confessions more generally, is epistemically puzzling. Through a comparison between Rivera's case and a canonical case from the epistemic injustice literature, this paper argues that the over-reliance in Rivera's case can be understood as an instance of testimonial injustice. The concept of testimonial injustice, which was first introduced by Miranda Fricker, refers to an injustice that a testifier suffers when her testimony is not judged by a listener solely based on its epistemic value. Instead, the listener improperly weighs the testimony as a result of a negative identity prejudice that she holds against the testifier.[149]

This paper aims to make two interconnected contributions to the literatures on epistemic injustice and criminal confessions, respectively.[150] First, through a comparison between Rivera's case and the fictional trial in *To Kill a Mockingbird*, I argue for a novel understanding of testimonial injustice that (a) accounts for the fictional trial case and the other canonical case, (b) clarifies how cases of credibility excess constitute testimonial injustice, and (c) establishes a previously-unrecognized aspect of testimonial injustice, which I term content sensitivity, that provides a unified account of the harm of testimonial injustice.

---

144 Loevy.
145 "DNA Exonerations in the United States."
146 A number of studies have shown that confessions, including false confessions, are assigned excessive weight by actual juries. For a case study, see Leo and Drizin, "The Problem of False Confessions in the Post-DNA World."
147 For a psychological study, see Kassin and Neumann, "On the Power of Confession Evidence."
148 See, e.g., Krishnamurthi, "The Case for the Abolition of Criminal Confessions."
149 For Fricker's full account of the theory, see Fricker, *Epistemic Injustice: Power & the Ethics of Knowing*.
150 I am not the first to observe a connection between false confessions and testimonial injustice. Jennifer Lackey argues that a false confession is an example of *agential* testimonial injustice, due to the denial of agency during the production of the testimony. In this paper, however, I set aside the issues regarding the production of the testimony and instead only consider the jurors' uptake of the testimony. For Lackey's argument, see Lackey, "False Confessions and Testimonial Injustice"; Lackey, "False Confessions and Subverted Agency"; Lackey, "False Confessions and Agential Testimonial Injustice."

Second, with this clarified understanding of testimonial injustice in place, the paper argues that criminal defendants—both those who are factually innocent and those who are factually guilty—may be vulnerable to testimonial injustices. By bringing together the literatures on testimonial injustice and criminal confessions, I demonstrate how Rivera's case can help clarify and refine the contours of testimonial injustice to create a clearer, more unified theory. In turn, the better understanding of testimonial injustice sheds light on an overlooked harm, namely the variety of epistemic injustices that defendants experience at and around trial.

The paper proceeds as follows: In §1, I use Fricker's canonical cases to clarify an important ambiguity regarding the types of prejudices involved in testimonial injustice. In §2, I demonstrate the parallels between Rivera's case and the fictional trial. In §3, I present the credibility excess debate and discuss the challenge of identifying cases with the right features. In §4, I explain how testimonial injustice is content-sensitive and demonstrate how content sensitivity reveals that the same harm occurs in cases of credibility deficit and in cases of credibility excess. In §5, I apply this clarified definition of testimonial injustice experienced by criminal defendants in trial. In §6, I outline some of the non-epistemic harms that result from these instances of testimonial injustice.

§1. Prejudice Type

Before we begin, let us address an important ambiguity regarding the types of prejudices involved in testimonial injustice. Fricker outlines various ways in which testimony may be improperly weighed. She focuses her definition, however, on cases in which a *negative identity prejudice* is the core of the testimonial injustice. Fricker defines negative identity prejudices as prejudices "against people owing to some feature of their social identity."[151]

She further restricts cases of testimonial injustice to instances where there is a *systematic*, as opposed to an *incidental*, negative identity prejudice at play. A systematic negative identity prejudice tracks a person across different dimensions of that person's life.[152] In contrast, incidental negative identity prejudices are highly localized. For example, Fricker considers a case in which a philosopher of science is primarily ignored at a large international science conference because her discipline is viewed as "out of touch."[153] While this case is an instance of a negative identity prejudice manifesting in the improper weighing of testimony, Fricker argues that this case and similar cases are not instances

[151] Fricker, *Epistemic Injustice: Power & the Ethics of Knowing*, 28.
[152] Fricker, 27. She defines these prejudices as "prejudices that 'track' the subject through different dimensions of social activity—economic, educational, professional, sexual, legal, political, religious, and so on."
[153] Fricker, 28.

of testimonial injustice because the prejudice is too limited. Thus, testimonial injustice can be preliminarily defined as follows:

> **Testimonial Injustice (TI)**: Testifier T suffers from a testimonial injustice if and only if T's testimony is improperly weighed by Listener L and L's improper weighing of T's testimony is a manifestation of a systematic negative identity prejudice that L holds against T.

So far, we have clarified that the identity prejudice must be both systematic and negative. In this section, I argue that there is another important clarification needed regarding the type of systematic negative identity prejudice that testimonial injustice captures, which is not made explicit in the literature. I then argue in favor of one of the possible interpretations in order to create a more consistent definition.

1.1 Prejudice Type Ambiguity

It is unclear whether the systematic negative identity prejudices involved in cases of testimonial injustice need to pertain to the testifier's capacity as a knower per se, or whether the systematic negative identity prejudices can pertain to a different aspect of the testifier's identity. Prejudices regarding a person's "capacity as a knower" are prejudices that pertain to a person's reliability as a transmitter of knowledge (i.e., the term includes prejudices about lying and deceiving). To demonstrate the difference between these categories, consider the following two examples of negative identity prejudices:

(1) Women are not smart.

(2) Women prefer domestic labor to non-domestic labor.

The first prejudice, that women are not smart, is a negative identity prejudice that pertains to a potential testifier's capacity as a knower. Suppose that the testifier is a woman, but the listener believes that women are not smart. In this situation, the listener may not believe the testifier as a result of the negative identity prejudice the listener holds.

The second prejudice, that women prefer domestic labor to non-domestic labor, does not directly pertain to a potential testifier's capacity as a knower. Suppose, however, that the testifier is a woman, and she states a preference for non-domestic labor. If the listener believes that women prefer domestic labor, then this negative identity prejudice may manifest in the listener not believing the woman's testimony. Even though the negative identity prejudice does not pertain to the woman's capacity as a knower per se, the prejudice can still interfere with the listener's weighing of the

testimony. As a result, the woman is epistemically harmed, since she is not taken seriously as a source of knowledge about her own experience.

The ambiguity regarding the type of prejudice leads to two divergent definitions of testimonial injustice:

> **TI (Narrow)**: Testifier T suffers from a testimonial injustice if and only if T's testimony is improperly weighed by Listener L and L's improper weighing of T's testimony is a manifestation of a systematic negative identity prejudice that L holds against T *regarding T's capacity as a knower*.
>
> **TI (Broad)**: Testifier T suffers from a testimonial injustice if and only if T's testimony is improperly weighed by Listener L and L's improper weighing of T's testimony is a manifestation of *any* systematic negative identity prejudice that L holds against T.

TI (Broad) would encompass all cases that fall under **TI (Narrow)**, but **TI (Narrow)** excludes some cases that fall under **TI (Broad)**.

Fricker never explicitly states whether she imagines that the prejudice must pertain to a person's capacity as a knower or if a more general prejudice could qualify. In her discussions of the two central examples of testimonial injustice, however, she explains the relevant prejudices as ones that pertain to the testifier's capacity as a knower.

The first canonical case is from Anthony Minghella's screenplay *The Talented Mr. Ripley*.[154] In the play, Marge Sherwood, a young woman, tries to share her (correct) belief that Tom Ripley killed her fiancé, Dickie Greenleaf, with Dickie's father, Herbert Greenleaf. Greenleaf responds, "Marge, there's female intuition, and then there are facts," thereby dismissing Marge's assertion that Tom Ripley, Dickie's supposed friend, had in fact killed Dickie.[155] As Fricker explains, Greenleaf's dismissal of Marge's testimony is a manifestation of a negative identity prejudice against women, namely that they are irrational and unreasonable.[156] The prejudice that women are irrational and unreasonable pertains directly to Marge's capacity as a knower. Since she is not rational, Greenleaf believes, he can discount or ignore her testimony. As such, the negative identity prejudice at play in the first canonical case is similar to prejudice (1), namely that women are not smart. Both prejudices pertain directly to a woman's capacity as a knower.

---

[154] Fricker, *Epistemic Injustice: Power & the Ethics of Knowing*, quoting Minghella, *The Talented Mr. Ripley*. The case is first mentioned on p. 9 and discussed throughout the book.
[155] Fricker, 9 quoting Minghella, *The Talented Mr. Ripley*, 130.
[156] For discussion of the case, see Fricker, *Epistemic Injustice: Power & the Ethics of Knowing*, 9–10, 15–16, 86–98.

The second central case of testimonial injustice comes from *To Kill a Mockingbird*.[157] Set in 1935, the novel tells the story of Tom Robinson, an American Black man who is on trial for raping a White woman, Mayella Ewell. Robinson attests to his innocence, and there is ample evidence of his innocence presented during the trial. Atticus Finch, Robinson's White defense lawyer, explains that the bruising pattern on Mayella must have come from a person who is left-handed. Atticus goes on to demonstrate that Robinson's left arm is disabled due to a childhood accident. Despite the fact that his innocence should be apparent to the jurors, Robinson is convicted.

According to Fricker's explanation of the case, the jurors discount Robinson's testimony due to the prejudice they hold against Black men. She notes that Robinson is convicted due to the jurors' perceptions of Robinson, which are "formed principally by the racial stereotypes of the day."[158] These racial stereotypes, she explains, are captured by Atticus' final speech to the jury. He tells the jurors to put aside the "assumptions—the evil assumptions—that *all* Negroes lie, that *all* Negroes are basically immoral beings, that *all* Negro men are not to be trusted around our women."[159] Atticus is clearly articulating the potential prejudices that he anticipates will interfere with the jurors' epistemic abilities to properly weigh evidence. Fricker goes on to explain that the negative identity prejudice "all Negroes lie" is the prejudice that manifests in the discounting of Robinson's testimony. Just as Greenleaf discounts Marge's testimony because women are irrational, the jurors discount Robinson's testimony because Black men are liars. As such, Fricker identifies a negative identity prejudice in the second central case that also pertains to the testifier's capacity as a knower.

In her own analysis of these two cases, Fricker seems to focus on prejudices that specifically pertain to the testifier's capacity as a knower. When Fricker later discusses the Robinson case throughout the book, she specifically refers to the prejudice about lying, without any mention of the other prejudices listed in Finch's speech.[160]

---

[157] Fricker, *Epistemic Injustice: Power & the Ethics of Knowing*. The case is first mentioned on p. 23 and discussed throughout the book. For discussion of this case in the literature, see, e.g., Wanderer, "Addressing Testimonial Injustice," 167–69; Pohlhaus, "Discerning the Primary Epistemic Harm in Cases of Testimonial Injustice," 100; McGlynn, "Epistemic Objectification as the Primary Harm of Testimonial Injustice," 161–63.

[158] Fricker, *Epistemic Injustice: Power & the Ethics of Knowing*, 25.

[159] Fricker, 25 quoting Lee, *To Kill a Mockingbird*, 208.

[160] Fricker's other mentions of the prejudice in Robinson's case include: "He will be perceived as a presumptuous, lying Negro" (23). "They cannot…perceive Tom Robinson as anything but a lying Negro" (36).

1.2 The Broad Definition

I argue, however, that the canonical cases are actually best understood as cases in which the relevant prejudices do not only pertain to the testifier's capacity as a knower. To account for these canonical cases, which are discussed throughout the literature after Fricker, the concept of testimonial injustice is best captured through **TI (Broad)**.

Let's start with Robinson's case. While the prejudicial belief that "*all* Negroes lie" clearly fits into **TI (Narrow)**, this specific prejudice alone cannot explain the outcome. Suppose that the jurors had only held the prejudicial belief that Black people lie frequently. If so, then the jurors would have discounted or completely disregarded Robinson's testimony. Yet, there had been other, stronger evidence presented in the trial from a source whom the jurors would have viewed as credible that pointed to Robinson's innocence. Atticus, a White man, presents evidence that a left-handed person attacked Mayella and that Robinson is not able to use his left hand. Even if the jurors had completely ignored Robinson's testimony, the jurors would have still needed to contend with the evidence from the non-Black lawyer that the attacker could not have been Robinson. Thus, if we only consider prejudices regarding the testifier's capacity as a knower, it is difficult to explain why Finch's testimony is also ignored, which leads to Robinson's conviction.

Moreover, if the jurors had only held the prejudice that "*all* Negroes lie," then one would expect racist members of Robinson's society to consistently ignore or discount Robinson's testimony on other matters. Yet, Fricker acknowledges that Robinson would likely have been believed in other contexts:

> Tom Robinson might have been relied on and trusted epistemically on certain matters even by the more thoroughly racist white citizens of Maycomb County—matters relating to his daily work, no doubt, and indeed many everyday matters of practical import.[161]

If the only prejudice at play is that "*all* Negroes lie," then one would expect that prejudice to play out consistently across different domains of Robinson's life. Since this is not the case, we need an additional explanation as to why he is believed in other contexts yet not believed in his criminal trial. Thus, relying exclusively on the prejudice regarding his capacity as a knower is insufficient to explain the other details of Robinson's life.

Instead, **TI (Broad)** can more fully explain the testimonial injustice in this case. In addition to the prejudice about Robinson's capacity as a knower, the jurors also hold a prejudice directly

[161] Fricker, *Epistemic Injustice: Power & the Ethics of Knowing*, 131. She explains the conflict between the jurors ignoring Robinson at trial and believing him in other contexts as the "tendency for incoherence in human prejudice" (131).

regarding the likelihood of Robinson's guilt. As Atticus Finch explains, the jurors come into the trial with the prejudicial belief that "*all* Negro men are not to be trusted around our women."[162] This prejudicial belief does not pertain to Robinson's capacity as a knower. Rather, this prejudicial belief leads the jurors to believe that Robinson had in fact committed the crime against Mayella, a White woman. Since the jurors possess a prejudicial belief regarding Robinson's guilt, any evidence to the contrary, even if presented by an otherwise credible source, may be ignored. Indeed, a hallmark of a prejudicial belief is that it is evidence-resistant, namely that people maintain their prejudicial belief even in the face of strong contrary evidence. Thus, a prejudicial belief about the likelihood of Robinson's guilt can more fully explain the canonical case. It explains the discounting of both Robinson's testimony about his innocence and Finch's testimony about the left-handedness of the actual perpetrator.

Furthermore, an underlying belief about Robinson's guilt best explains why his testimony is ignored in one context and relied upon in different contexts. Outside of the courtroom, the prejudicial belief pertaining to Robinson's guilt is not relevant. When Robinson is consulted regarding his daily work, the racist White members of Robinson's society may trust Robinson on those matters, since their prejudicial belief about Black men around White women is not relevant. Thus, Robinson's case of testimonial injustice can be fully explained by their possession of a prejudicial belief that Robinson likely committed the crime against a White woman because he is a Black man.

In Marge's case, the negative identity prejudice that women are irrational seems to pertain only to Marge's capacity as a knower. Yet, that formulation of the prejudice does not fully capture why the testimonial injustice occurs. The primary reason that Marge believes Tom killed Dickie is that she finds the ring she gave to Dickie in Tom's apartment. She tells Greenleaf that Dickie had promised to never remove it. Tom, who had in fact killed Dickie, tells Greenleaf that Dickie took off the ring because he was unfaithful to Marge. As a result, Greenleaf hears Marge's testimony as "a sweetheart's wishful thinking."[163] The negative identity prejudice at play is not just that women are irrational in general but rather that women are irrational about certain topics. Greenleaf discount's Marge's testimony in this instance because he believes that women are irrational and ignorant when it comes to romantic relationships, love, and male fidelity. Thus, in Marge's case of testimonial injustice, the negative identity prejudice pertains to her capacity as a knower *about certain topics*.

---

[162] Lee, *To Kill a Mockingbird*, 208.
[163] Fricker, *Epistemic Injustice: Power & the Ethics of Knowing*, 87.

A more fine-grained understanding of the content of the prejudice demonstrates that prejudices that appear to be about the testifier's capacity as a knower may actually be grounded in a more general prejudice that is not only about the testifier's capacity as a knower. In Marge's case, for instance, the prejudice that women are irrational about love is likely rooted in a more general prejudice about women, such as the belief that women only care about being loved by a man. Similarly, the prejudice that seems to pertain directly to Robinson's capacity as a knower, namely that Black men lie, is likely also rooted in a more general prejudice. As demonstrated by the fact that Robinson would have been believed in other contexts, the prejudicial belief that Black men lie should be understood in a more specific way: Black men lie about certain topics, like their guilt, but not other topics, like their work and daily matters. The differentiation between guilt and work reveals that the prejudice is rooted in a more general prejudice about Black men and how they act, such as the prejudicial belief that Black men "cannot be trusted around our women." Thus, even when the prejudice appears to be about a person's capacity as a knower, it may still be rooted in a more general prejudice.

TI (Narrow) struggles to fully capture Greenleaf's dismissal of Marge's testimony and the jurors' verdict in Robinson's case, while **TI (Broad)** fully explains both cases. Furthermore, since many scholars after Fricker draw on these two canonical cases, it makes sense to favor a definition that can properly account for the central cases in the literature. Now that we have clarified the relevant prejudices, let us turn to our central issue: the comparison between Robinson's case and Rivera's case.

§2. Canonical Case Comparison

Now that we have clarified the definition and the prejudices involved in Robinson's case, we are prepared for the central comparison. As discussed, **TI (Narrow)** only picks up the prejudice that Black men lie generally. **TI (Broad)**, however, helps demonstrate that the jurors believe that, since Robinson is a Black man, he likely committed the crime and subsequently lied about it.

Now, let us turn to the comparison between his case and Rivera's case. Though Robinson's case is fictional and Rivera's real, both defendants' respective testimonies are improperly weighed. In Rivera's case, there is ample evidence that indicates that he is innocent. Since the DNA evidence found did not match Rivera's DNA, the prosecution came up with a highly implausible theory of the case.[164] Thus, the heavy reliance on Rivera's confession *despite* the evidence pointing to his innocence seems

[164] To add to the implausibility, the police had records demonstrating that Rivera was over two miles away from the crime scene when the crime occurred, since he was under electronic monitoring at the time. The theory presented was that Rivera slipped out of the ankle monitor, traveled two miles to commit the crime, and then put the monitor back on—all without being detected.

unreasonable. If an unbiased listener were presented with the evidence in the case, the listener would likely conclude that Rivera is innocent. At the very least, the listener would realize that there is a *reasonable doubt* that Rivera could be innocent. Thus, the jurors' conclusion that Rivera's confession is indeed true and reliable despite all the evidence to the contrary demonstrates that the jurors place excessive weight on Rivera's testimony.

Similarly, in Robinson's case, there is plenty of evidence that indicates he is innocent. Finch's demonstration in court that Robinson's arm was disabled from a childhood accident alone should have proved that Robinson would not have been capable of inflicting the bruising pattern on Mayella. An unbiased juror presented with that evidence, along with Robinson's claim of innocence, would have likely adopted Robinson's testimony and subsequently voted to acquit him. Just like Rivera's case, at the very least an unbiased juror would have a reasonable doubt regarding Robinson's innocence.

In both cases, the improper weighing of the evidence is the manifestation of a negative identity prejudice against the defendant. In Robinson's case, the prejudice is clearly captured by Finch's closing speech to the jury. The jurors possess the prejudicial belief that Black men "cannot be trusted around our women." As such, the belief pertains directly to Robinson's alleged guilt, since he is accused of raping a White woman.

There are potentially two related negative stereotypes in Rivera's case that could explain why the jurors ascribe excessive credibility to his confession. The first negative identity prejudice is one ascribed to all criminal defendants, and the second negative identity prejudice is one ascribed to Latino or Hispanic men in the US.

First, criminal defendants are often viewed as guilty, regardless of whether or not they have already been convicted. People who are accused of crimes are often detained before trial because they are considered to be "dangerous."[165] One of the implicit justifications for pre-trial detention is the assumption that someone who is accused of a crime likely did, in fact, commit the crime.[166] If the state relies on such an assumption, it communicates to others that those who are accused of crimes, namely criminal defendants, likely committed the crimes.[167] Furthermore, scholars have argued that acquitted

[165] Mayson, "Dangerous Defendants."
[166] Mayson.
[167] This belief may in fact be true. In the US, for example, the vast majority of federal criminal defendants are found guilty. See Gramlich, "Only 2% of Federal Criminal Defendants Go to Trial, and Most Who Do Are Found Guilty," 2.

defendants should receive compensation for the reputational harm of being a defendant.[168] Since these defendants are acquitted, it appears that the reputational harm is a result of people's pre-conceived notions that those who are accused of committing a crime did, in fact, commit the crime.

One may be concerned, however, that this prejudice is too temporary and constrained to qualify as testimonial injustice. Testimonial injustice specifically is meant to only capture systematic prejudices and exclude incidental prejudices, such as the prejudice against the philosopher of science at the conference. Thus, the prejudice held against the testifier needs to be relevant in multiple places and at different times in a person's life.[169] If the prejudice only affects people in court while they are criminal defendants, it seems that it would be a case of incidental negative identity prejudice.

I respond, however, that the discrimination and prejudice experienced by criminal defendants transcend both the court room and the person's time as a criminal defendant. People face employment discrimination as a result of being charged with—but not convicted of—a crime.[170] This employment discrimination occurs beyond the bounds of the court and often long after a person is no longer technically a criminal defendant. Thus, it appears that a negative identity prejudice against criminal defendants would be sufficiently lasting, thus meeting the systematic requirement.

There may be a second, additional prejudice at play in this case since Rivera is a Hispanic man. Social scientists Stephen Demuth and Darrell Steffensmeier have found that Hispanic men are more likely to be convicted of a crime and are more likely to received harsher sentences than are White men.[171] Moreover, studies that analyze court transcripts from criminal cases have found extensive bias against Mexican Americans in US criminal trials.[172] Thus, if the jurors possess the racist prejudice that Hispanic men often commit crimes, the jurors may also be more inclined to over-weigh Rivera's confession, since such testimony supports their racist views.

The prejudices at play in Rivera's case mirror those in Robinson's fictional case. Like Rivera, Robinson is a criminal defendant. If there is a negative identity prejudice against criminal defendants, regardless of race, it would likely apply in both cases. Moreover, the anti-Black prejudice that

---

168 See, e.g., Burnett, Hoyle, and Speechley, "The Context and Impact of Being Wrongly Accused of Abuse in Occupations of Trust"; Friedman, "Destigmatizing the Innocent Defendant: Rehabilitating Reputation After an Assault Under Color of Law."

169 Fricker, *Epistemic Injustice: Power & the Ethics of Knowing*, 27–29.

170 Geffen, "The Collateral Consequences of Acquittal."

171 Demuth and Steffensmeier, "Ethnicity Effects on Sentence Outcomes in Large Urban Courts."

172 Espinoza and Willis-Esqueda, "Defendant and Defense Attorney Characteristics and Their Effects on Juror Decision Making and Prejudice against Mexican Americans." Moreover, former US President Donald Trump's racist rhetoric drew a connection between Latino men and sexual violence. See, e.g., Gonzalez, "Stereotypical Depictions of Latino Criminality."

Robinson's jury possesses is similar to an anti-Hispanic prejudice that Rivera's jury may possess. Both of these prejudices are about the criminality of members of a group. In both cases, jurors likely possessed these negative identity prejudices before even hearing the evidence.

If the jurors possess the negative identity prejudice that criminal defendants are likely guilty of the crime for which they are accused, this prejudicial belief would explain the over-weighing of Rivera's false confession. If the jurors had entered the trial with the belief that Rivera, a Hispanic man, is likely guilty, then it makes sense that such a prejudice would lead them to overly rely on testimony that appears to confirm their prejudice. Similarly, if the jurors in *To Kill a Mockingbird* entered the trial with the belief that Robinson is likely guilty, then it makes sense that such a prejudice would lead them to discount testimony that appears to undermine that conclusion. Thus, both juries' guilty verdicts can be explained by the jurors' possession of a prejudicial belief regarding the likelihood that the defendant committed the crime.

Given all of these parallels, Rivera's case appears to also be an instance of testimonial injustice. His case meets the **TI (Broad)** conception of testimonial injustice, since (a) Rivera's testimony is improperly weighed by the jurors, and (b) the improper weighing of the testimony is a manifestation of a systematic negative identity prejudice that the jurors hold against Rivera. Thus, we can understand the jurors' epistemically puzzling behavior in Rivera's case.

One may object, however, that Rivera's case is importantly different from Robinson's. Rivera testifies against his own interests, while Robinson testifies in accordance with his own interests. Perhaps, then, the jurors have good reason to take Rivera's testimony especially seriously. In fact, it may seem that the jurors may have placed excessive weight on Rivera's testimony solely because it is clearly against Rivera's interest to confess. If this is the only explanation for the over-weighing of Rivera's testimony, then it would not constitute an instance of testimonial injustice, since it would not be a manifestation of a negative identity prejudice.[173]

I respond, however, that there still is a limit to which it is reasonable to accept someone's testimony in the face of far more compelling contrary evidence. It may be the case that the jurors do have good reason to take Rivera's testimony more seriously given that it was contrary to his own interests. Indeed, it makes sense that Rivera's confession would be taken more seriously than

[173] If this belief is prevalent, it calls for better education of jurors regarding the rate of false confessions and the psychological dynamics that give rise to false confessions. For a study of jurors' beliefs about confessions, see Costanzo, Shaked-Schroer, and Vinson, "Juror Beliefs About Police Interrogations, False Confessions, and Expert Testimony." For a study regarding how jurors react to information about the reliability of confessions, see O'Donnell and Safer, "Jury Instructions and Mock-Juror Sensitivity to Confession Evidence in a Simulated Criminal Case."

testimony from a random witness, or from someone who may have benefited from providing the testimony. Yet, despite the fact that counter-self-interest testimony may appear to be reliable, it is not an infallible source of testimony. Rather, an additional belief or prejudice is required in order to explain why the jurors believe Rivera's testimony despite the mountain of evidence pointing to the contrary. Moreover, given the fact that Rivera's testimony works against his own interests, one would perhaps expect the jurors to be more skeptical of the circumstances under which Rivera had provided the testimony.[174]

Instead, the primary difference between these cases is that Rivera's case is one of credibility excess, while Robinson's case is one of credibility deficit. But, whether cases of credibility excess should be included in the concept of testimonial injustice is debated. In the coming sections, I lay out the credibility excess debate and situate Rivera's case within that debate.

§3. Credibility Excess Debate

Fricker originally laid out two potential categories of testimonial injustice: credibility deficit and credibility excess.[175] Fricker then argued that credibility excess cases are not instances of testimonial injustice. She considers a case of an over-burdened doctor whose patients ask him specialized questions.[176] The doctor does not have the necessary specialty training needed to answer the questions, but the patients do not have access to any other medical professionals, so the doctor must answer their questions. In such a case, the excessive reliance on the doctor's testimony may be harmful. Yet, Fricker argues that the harms experienced are non-epistemic harms, since the over-reliance places an undue burden on the doctor to advise the patients correctly. Fricker ultimately argues that testimonial injustice requires that the testifier be harmed in her *capacity as a knower*, and that cases of credibility excess will not meet this description.[177]

Indeed, if Fricker did in fact favor a **TI (Narrow)** understanding, it is clear why she would exclude instances of credibility excess. If the concept of testimonial injustice only includes negative identity prejudices regarding the testifier's capacity as a knower, then it would be difficult to construct

[174] Indeed, Jennifer Lackey argues that there is an epistemic harm that occurs when testimony is extracted against a person's own interest, thereby denying epistemic agency. See Lackey, *Criminal Testimonial Injustice*.
[175] Federico Luzzi argues that due credibility could also be testimonial injustice. He contends that a testifier could have their testimony properly weighed because one negative identity prejudice leads to the over-weighing of the testimony, while a second negative identity prejudice regarding the same testifier can lead to the under-weighing of testimony. In order for such a possibility to exist, credibility excess as testimonial injustice would first need to be established. As such, the arguments presented in this paper bear on Luzzi's view. See Luzzi, "Testimonial Injustice Without Credibility Deficit (or Excess)."
[176] Fricker, *Epistemic Injustice: Power & the Ethics of Knowing*, 18.
[177] Fricker, 21.

a case in which a listener holds a negative identity prejudice regarding the testifier's capacity as a knower and still assigns the testimony excessive weight.

Since Fricker made the claim that testimonial injustice only includes instances of credibility deficit, a variety of scholars have attempted to identify cases of testimonial injustice in which the testifier is afforded credibility excess.[178] In order to understand the debate, it is important to clarify that none of the scholars are claiming that cases of credibility excess fit Fricker's definition of testimonial injustice, since Fricker explicitly excludes such cases. Rather, scholars are arguing in favor of modifying the concept of testimonial injustice to include cases of credibility excess for different reasons.

There are two primary ways of identifying instances of credibility excess as testimonial injustice in the literature. First, some scholars, such as José Medina, reframe credibility on a larger scale.[179] Instead of assessing credibility relative to the deserved credibility of the testifier in that moment, Medina argues that credibility is proportionally distributed across testifiers, based on how social identities intersect in a given situation.[180] Thus, every time there is a credibility deficit afforded to one testifier, there is necessarily a credibility excess afforded to another testifier. For instance, Medina argues that in Robinson's trial, it is impossible to view Robinson's credibility deficit without also seeing the excess credibility afforded to Mayella or to the other White people involved in the trial.[181]

The second route taken by some scholars is to attempt to identify cases in which people experience both epistemic harms and excess credibility.[182] For instance, Emmalon Davis argues that positive stereotyping can lead to excess credibility placed on marginalized knowers, which in turn creates epistemic burdens.[183] She explains the problem of compulsory representation, namely that marginalized knowers are put into positions in which they are expected to speak on behalf of their identity group. She also coins the terms *epistemic exploitation*, describing the phenomenon of placing marginalized knowers in positions in which they are required to "educate" others.

The cases of excess credibility presented by Medina and Davis, however, do not demonstrate the same type of harm found in canonical cases of credibility deficit. In the canonical cases, the

---

[178] See, e.g., Medina, "The Relevance of Credibility Excess in a Proportional View of Epistemic Injustice"; Davis, "Typecasts, Tokens, and Spokespersons."
[179] For another scholar who takes a similar approach, see, e.g., Yap, "Credibility Excess and the Social Imaginary in Cases of Sexual Assault."
[180] Medina, "The Relevance of Credibility Excess in a Proportional View of Epistemic Injustice."
[181] Medina.
[182] Davis, "Typecasts, Tokens, and Spokespersons."
[183] Davis.

improper weighing of the testimony is a manifestation of a negative identity prejudice against the testifier herself. In Medina's cases, the people experiencing the excess credibility are not harmed by the excess credibility. Instead, the excess credibility is part of a larger injustice that harms those who experience the relative credibility deficit. In Davis' examples, the harms surround the improper weighing of testimony but are not themselves the manifestation of a prejudice. Rather, the marginalized knowers are burdened before testifying when they are placed in positions in which they are expected to educate others. Epistemic exploitation can be a time and an emotional burden. Though these harms are important, they are not the same type of harm involved in canonical testimonial injustice cases. Moreover, Davis notes that the core of the problem is actually a credibility deficit, since the marginalized knowers are invited into the knowledge exchange *only* because of their identity (e.g., gender, race, religion). Otherwise, they would have been excluded. Thus, both Medina and Davis do not provide cases in which a person experiences credibility excess and the excess is itself the primary epistemic harm.

Though scholars like Medina and Davis have productively explored how credibility excess functions, there has yet to be a credibility excess case in the literature with the same type of harm found in canonical cases of credibility deficit testimonial injustice. In the coming section, I demonstrate how testifiers who experience credibility excess and testifiers who experience credibility deficit can be epistemically harmed in the same way.

§4. Content Sensitivity

Thus far, I have demonstrated how Robinson and Rivera's cases are similar in many respects. The primary difference is that Robinson experiences a credibility deficit, while Rivera experiences a credibility excess.

The pair of parallel cases helps us better understand what I term the content-sensitive nature of testimonial injustice.[184] The content-sensitive nature of testimonial injustice not only explains how credibility is assigned but also how testifiers are similarly epistemically harmed, regardless of whether they receive a credibility excess or deficit.

[184] This concept is distinct from Dembroff and Whitcomb's Content-Focused Injustice (CFI), which applies when the testifier's identity and the testimony content come apart. Finch's testimony, for example, would be an example of CFI, since he is not a Black defendant, yet his testimony is ignored due to its content. My concept of content sensitivity, however, considers how both the content *and* the testifier's identity together create testimonial injustice, and thus excludes cases like Finch's testimony. See, e.g., Dembroff and Whitcomb, "Content Focused Epistemic Injustice."

In both Rivera's case and Robinson's case, their respective testimonies are improperly weighed. Robinson experiences a credibility deficit, while Rivera experiences a credibility excess. But the credibility excess or deficit reflects how their respective testimonies interact with the prejudice held by each jury. In both cases, the jurors are inclined to believe that the defendant is guilty. Since Robinson testifies that he is innocent, Robinson's testimony goes against the jurors' preconceived belief in his guilt. The jurors therefore discount Robinson's testimony. Rivera's confession, however, confirms the jurors' preconceived belief in his guilt. Thus, the jurors overly rely upon Rivera's testimony. Whether or not someone experiences a deficit or excess does not only depend on the prejudices held against the testifier. Rather, the credibility (or lack thereof) received is closely tied to whether the testimony confirms or undermines the negative identity prejudice that the listener holds. The *content* of the testimony, therefore, plays a role in determining whether the listener discounts or overcounts the testimony.

The content-sensitive nature of testimonial injustice captures the way in which testimonial injustice is epistemically harmful in both cases of credibility excess and cases of credibility deficit. Fricker emphasizes that the harm that arises in cases of testimonial injustice is not only a practical harm but rather that the testifier is "wronged specifically in her capacity as a knower."[185] Thus, the relevant harm of testimonial injustice is an *epistemic* harm; the injustice is that listeners "undermine, insult, or otherwise withhold proper respect for the speaker *qua* subject of knowledge."[186] In Rivera's case, for example, the jurors actually disrespect Rivera by eagerly adopting his false testimony. A good epistemic agent engaging in a knowledge exchange would critically evaluate testimony and only adopt testimony that seems correct and plausible.[187] Instead, the jurors do not listen to Rivera's testimony with an open mind, nor do they subject the testimony to critical reflection. If they had, they would have realized that the confession was likely false. Instead, Rivera's testimony is subverted by the jurors to confirm a prejudice that the jurors already held against Rivera. By not evaluating the testimony, the jurors demonstrate a lack of respect for Rivera as a knower.[188]

---

[185] Fricker, *Epistemic Injustice: Power & the Ethics of Knowing*, 20. Fricker acknowledges that people may suffer practical harms, including wrongful convictions, as a result of an instance of testimonial injustice.

[186] Fricker, 20.

[187] One may be concerned that my view entails that uncritically adopting the testimony of a woman who says that she was sexually assaulted constitutes testimonial injustice. In such cases, however, the uncritical adoption of the testimony is not a manifestation of a *negative* identity prejudice against the woman herself (though it may be a manifestation of a negative identity prejudice against someone else or men more generally).

[188] Notably, a person can be disrespected as a knower, even when she is not sharing knowledge, as in Rivera's case. Testimonial injustice does not require knowledge or truthful testimony. Rather, it is about the listener's attitude and the disrespect shown to the listener.

Similarly, in Robinson's case, the jurors do not show respect for Robinson's capacity as a knower. Instead of listening to and engaging with Robinson's testimony, the jurors seem to already have their minds made up regarding Robinson's guilt. Thus, Robinson's testimony is quickly rejected by the jurors, since it appears to undermine the pre-existing prejudice that Robinson is likely guilty. By rejecting the testimony regardless of the strength of the evidence, the jurors demonstrate a lack of respect for Robinson as an agent who is attempting to communicate knowledge. Thus, in both cases, Robinson and Rivera suffer epistemic harms.

People are not only harmed when they experience a credibility deficit. Rather, people are also harmed when their testimony is adopted only because it appears to support a prejudice that the listener holds. In this way, we can see how people are objectified in all cases of testimonial injustice. Instead of being treated as a source of knowledge, the person's testimony is merely filtered through the listener's bias: if it supports the prejudice, it is adopted, and if it undermines the prejudice, it is rejected. Either way, the listener emerges from the supposed knowledge exchange with the same beliefs. As a result, even if excessive weight is placed on the testimony, the testifier is not genuinely treated as a source of knowledge.

Rivera's case helps us see how other cases of credibility excess fit the same structure and thus may also be cases of testimonial injustice. In this case, Rivera's testimony confirms jurors' pre-existing prejudicial belief, which leads them to assign the testimony extra weight. (In Rivera's case, we can most clearly see that the testimony is improperly weighed, since it is clearly false in the context of all the other evidence.) Yet, any time that a listener's prejudice against a testifier leads the listener to over-eagerly adopt testimony that confirms that prejudice, the testifier experiences a credibility excess testimonial injustice.

This understanding of the content-sensitive nature of testimonial injustice fits into Fricker's more general view of epistemic objectification. She draws an analogy between epistemic objectification and sexual objectification. Fricker discusses Catharine MacKinnon's argument regarding the objectification of women—when a woman says "no" to a sexual advance by a man, she is ignored because the man views her as an object. According to MacKinnon, the objectification of women is constructed by pornography, which teaches men that women are subordinate and desire submission. As a result, when a woman says "no," the man hears the rejection as part of a taught narrative in which women say "no" as a part of a "script" in which they actually desire the sexual advances.

A similar objectification process may play out in the courtroom. According to MacKinnon, men are taught a narrative that women, regardless of whether they say "no" or "yes," always desire

sexual encounters. Similarly, it is conceivable that jurors may believe that criminal defendants, regardless of whether they testify to their guilt or innocence, are guilty. As such, the defendant's testimony functions similarly to that of the woman rejecting the sexual encounter. When a woman says "no," the man may think: the woman actually desires sex, she is just saying "no" as part of the script. Similarly, if the defendant says he is innocent, the jurors may think: the defendant is actually guilty, he is just saying that he is innocent as part of the script. The same applies to the woman who says "yes" or the criminal defendant who confesses. In both cases, the testimony aligns with the *social script* expected by the listener. As such, the woman and the defendant are not truly being heard. Rather, they are viewed as objects and their testimony is not taken seriously.

Sexual objectification and epistemic objectification are similar types of harm. In both instances, the person is not treated as a person but rather as an object, available to be used by another person. At the core of Fricker's argument regarding the harm of testimonial injustice is this concern regarding epistemic objectification: people are treated as objects carrying testimony, which can either be adopted or disregarded, as opposed to knowers who are engaging in a knowledge exchange.

To review, we began with an ambiguity about the types of prejudices involved in testimonial injustice. **TI (Broad)** best accounted for the canonical cases discussed by Fricker and throughout the literature. With the clarified definition, we established the parallels between Rivera's case and Robinson's case. The primary difference between the cases was the difference between credibility excess and credibility deficit. By introducing the notion of content sensitivity, I showed that Rivera and Robinson are harmed in the same way, even if one experiences a credibility excess and one a credibility deficit. The content of the testimony, and how it corresponds to the prejudice, determines whether the testimony is quickly dismissed or eagerly adopted. Either way, the testifier is objectified and disrespected, thus providing a unified account of the harms of testimonial injustice.

Notably, none of these arguments work with the **TI (Narrow)** definition. According to that view, Rivera's case would not constitute testimonial injustice, since the jurors possess a general prejudice about his guilt, not a prejudice exclusively about his capacity as a knower. Moreover, cases of credibility excess would not be possible, since it is difficult to have a case in which a *negative* identity prejudice about a person's capacity as a knower leads the listener to give *excess* weight to the testimony. Finally, without the ability to compare cases of credibility excess and deficit, it is more difficult to see how the content of the testimony interacts with the prejudice. **TI (Narrow)** would not be content sensitive since the prejudice, if exclusively about a person's capacity as a knower, would lead the

listener to discount everything that the testifier says, regardless of the topic or content of the testimony.

In the coming section, I apply this clarified understanding of testimonial injustice to the criminal trial arena.

§5. Criminal Trials

Now that content sensitivity has provided a clarified understanding of the harm of testimonial injustice, let us turn to the question of how criminal defendants experience epistemic harm at and around trial. I first explain how testimonial injustice manifests and then demonstrate how silencing—another form of epistemic harm—similarly occurs at and around trial.

As discussed previously, people may hold a general prejudice against criminal defendants. If the jurors possess a strong prejudicial belief in a defendant's guilt, then the defendant is likely to experience testimonial injustice, regardless of whether the defendant states her innocence or guilt. If the defendant testifies that she is innocent, her testimony may be over-eagerly rejected as a result of the pre-existing prejudicial belief, as is the case with Robinson's testimony. Conversely, if the defendant testifies that she is guilty, her testimony may be over-eagerly adopted as a result of a pre-existing prejudicial belief, as is the case with Rivera's testimony. In either scenario, the testifier is not treated as a true source of knowledge, but rather as a tool to affirm the listener's prejudice.

Importantly, testimonial injustice can also occur when the defendant is factually guilty. To demonstrate how such a case could work, let us first consider a case without testimonial injustice. Suppose that a defendant has truthfully confessed to committing a crime, but there is other evidence in the case that casts doubt on whether the defendant indeed committed the crime. If the jurors did not have pre-existing prejudicial beliefs, the jurors would weigh all of the evidence together, recognizing that the confession appears to conflict with some of the other evidence in the case. Even if the jury ultimately convicted the defendant, the defendant does not experience a testimonial injustice, since the jurors treat the defendant and her testimony as a source of knowledge.

Now consider a similar case, except this time, the jurors possess a negative identity prejudice against the defendant. Before the jurors heard any evidence in the case, they assume that the defendant is guilty. Instead of thoughtfully weighing all of the evidence, the jurors immediately adopt the confession and do not think twice about the contradictory evidence. In such a scenario, the jurors place excessive weight on the defendant's true testimony. Yet, the defendant still suffers from a testimonial injustice. Her testimony is given excessive weight merely because it confirms the pre-

existing prejudice. In this way, the defendant is not treated as a source of knowledge, but rather her testimony is used to affirm the listeners' pre-existing prejudices against her.

Thus, regardless of whether the defendant is factually innocent or guilty, and regardless of whether the testimony is true or false, testimonial injustice can occur if jurors hold pre-existing prejudicial beliefs against defendants. The testimony is adopted or rejected based on the content of the testimony—whether it confirms or undermines the pre-existing prejudice. In all scenarios, the testifier is not treated as a real source of knowledge and is thus epistemically harmed.

If, in fact, this picture of the testimonial injustices at trial is correct, then it is unsurprising that a different but related epistemic harm is also occurring. Fricker explains that widespread testimonial injustice often coincides with *silencing*. Though silencing and testimonial injustice often come together, they are distinct forms of epistemic harm.

Fricker explains that two types of silencing arise from widespread testimonial injustice.[189] The first type of silencing is pre-emptive—the speaker is silenced by never testifying in the first place. In the US, for instance, criminal defendants rarely testify at their own trials.[190] Approximately 98% of criminal defendants never go to trial in the US.[191] Of those who do go to trial, it is estimated that approximately 50% of criminal defendants do not testify at their trials.[192] Indeed, criminal defendants even opt to not testify when they are innocent. Approximately 40% of people who were exonerated with DNA evidence in the US declined to testify in their original trials, likely because they were advised that they would not be believed by the jury.[193] Thus, pre-emptive silencing appears to occur to criminal defendants as well.

The second type of silencing described by Fricker is a silencing that arises from objectification. In objectification-based silencing, the person does attempt to testify, but the lack of uptake is itself silencing. The idea that "uptake" is required for communication to occur comes from Jennifer Hornsby and Rae Langton's interpretation of J. L. Austin's speech act theory.[194] They argue that a lack of hearer uptake is itself silencing, since uptake is a necessary condition for communication. If the testifier is prevented from communicating, then the testifier is silenced even if she makes auditory

---

189 Fricker, *Epistemic Injustice: Power & the Ethics of Knowing*, 129–42.
190 Natapoff, "Speechless."
191 Subramanian et al., "In the Shadows: A Review of the Research on Plea Bargaining."
192 Schulhofer, "Some Kind Words for the Privilege against Self-Incrimination the Bill of Rights Yesterday and Today," 329–30.
193 Blume, "The Dilemma of the Criminal Defendant with a Prior Record-Lessons from the Wrongfully Convicted," 489.
194 Hornsby and Langton, "Free Speech and Illocution." For Fricker's discussion of this view, see Fricker, *Epistemic Injustice: Power & the Ethics of Knowing*, 140–42.

sounds. In the context of sexual objectification, the lack of uptake occurs when a woman's "no" is ignored. The lack of uptake by the man is a form of silencing, even if the woman speaks. Similarly, in criminal trials, the lack of uptake by the jurors of the defendant's testimony can be a form of silencing. Even if the defendant testifies, she may experience silencing if there is no uptake of her testimony. Thus, the objectification-based silencing can also be observed in the context of criminal defendants.

These two types of silencing show how additional epistemic harms can exist. If the defendant does not testify, the defendant may experience pre-emptive silencing. Even if the defendant does testify, the defendant similarly may experience objectification-based silencing if there is no uptake. Even if there is uptake, a negative identity prejudice may manifest in the jurors' improper weighing of the defendant's testimony. Thus, defendants may be epistemically harmed through testimonial injustice, silencing, or both.

§6. Non-Epistemic Harms

In the previous section, I considered the potential epistemic harms—both misplaced credibility and silencing—experienced by criminal defendants at trial. I will now briefly turn to the question of non-epistemic harms that result from the epistemic harms discussed above.

First and foremost, there is the obvious concern regarding wrongful convictions. If jurors enter criminal trials with a prejudicial belief against criminal defendants, the jury is more likely to find the defendant guilty, even in cases in which the defendant is innocent. The testimonial injustice Rivera experienced, for instance, contributed to his conviction, which in turn led to Rivera spending 20 years incarcerated for a crime he did not commit. Thus, innocent criminal defendants may experience non-epistemic harms as a result of the testimonial injustices they experience at trial.

Second, testimonial injustice can undermine the fairness of a given trial, even if a person is factually guilty. For instance, a person may be convicted when there is insufficient evidence for conviction. Recall the case in which a guilty defendant confesses, yet there is other evidence in the case that casts doubt on the defendant's guilt. If there is insufficient evidence for conviction, then a conviction based on a prejudicial belief in the defendant's guilt seems highly problematic. Moreover, it seems as though there is an element of luck involved. Due to the prejudicial belief, that same jury would have convicted in a variety of scenarios, including ones in which the defendant was innocent. Thus, it is more of a coincidence that the defendant is in fact guilty. Convictions on the basis of prejudicial beliefs and luck seem highly problematic.

Third, testimonial injustice can undermine fairness across trials if it leads to the conviction or acquittal of certain groups at higher or lower rates. In Robinson's fictional case, for instance, we showed that the jurors likely hold the prejudicial belief that "*all* Negro men are not to be trusted around our women." Similarly, in Rivera's case, it is possible that the jurors possessed a prejudicial belief that Hispanic men are more likely to commit crimes. These negative identity prejudices clearly intersect with the question of the likelihood of guilt. Thus, if the prejudice leads the jurors to believe that the person is likely to be guilty, then Black and Hispanic defendants may be wrongfully convicted at a higher rate. Conversely, suppose that jurors hold the prejudicial belief that women are docile and thus unlikely to commit crimes. If so, then women would likely be acquitted at a disproportionately high rate.[195] In both scenarios, the outcome of a trial is affected by jurors' prejudicial beliefs regarding the identity of the defendant. As such, the overall fairness of the system is called into question if jurors' prejudicial beliefs about defendants' identities affect the likelihood of conviction.

Finally, if criminal defendants believe that they are likely to face a prejudicial jury, this belief affects the stakes of plea bargaining, regardless of whether or not the defendant is factually guilty. This effect occurs in multiple stages: (1) if a defendant believes that jurors may possess negative identity prejudices, then the defendant may believe that there is an increased chance of conviction at trial. (2) If the defendant believes that conviction at trial is likely, then the defendant may be more likely to take a plea bargain. (3) If the prosecutor knows that the defendant believes that conviction at trial is likely, then the prosecutor may expect that the defendant will accept a less desirable plea bargain. (4) If the prosecutor expects that the defendant will accept a less desirable plea bargain, then the prosecutor may offer a less desirable plea bargain. (5) If the defendant is offered a less desirable plea bargain, then the defendant may end up serving a longer sentence than the defendant would have otherwise served.

To demonstrate how the perception of prejudicial juries leads to longer sentences, consider the following example. Suppose that a defendant believes that she will face a fair jury and that there is a 50% chance of conviction at trial. If convicted, she will face a 25-year sentence. Alternatively, she can accept a plea deal and receive a 15-year sentence. In this scenario, the defendant is choosing between a 50% chance of receiving a 25-year sentence and a 100% chance of receiving a 15-year sentence. Alternatively, suppose that the defendant believes that she will face a prejudicial jury and that there is a 95% chance of conviction at trial. Given that perceived high likelihood of conviction,

[195] Indeed, women are incorrectly acquitted at a higher rate than are men in the US. See Strub and McKimmie, "Sugar and Spice and All Things Nice."

the prosecutor instead offers a plea deal in which she would receive a 20-year sentence. In this alternative scenario, the defendant is choosing between a 95% chance of 25 years and a 100% chance of 20 years. Given these odds, the defendant likely would choose the 100% chance of 20 years. The 20-year sentence, however, is five years longer than the 15-year plea deal that would have been offered had the defendant seriously considered going to trial. As such, just the *perception* of prejudicial juries may lead to longer sentences across the board. For this reason, even factually innocent defendants may be more inclined to take a plea deal over risking a longer sentence following a trial conviction.[196]

Given the cascade of non-epistemic harms created by testimonial injustice at trials, we see that defendants can be affected by these epistemic injustices in many ways. Those who go to trial may experience testimonial injustice or silencing. Those who do not go to trial may be offered and accept worse plea deals if it looks like trial outcomes are not only based on evidence. Thus, *all* defendants—regardless of guilt or innocence, confession or no confession, true confession or false confession, testimony or silence, opting for trial or a plea bargain—are implicated when jurors bring their prejudicial beliefs to trials.

Conclusion

In this paper, I brought together literature on testimonial injustice and criminal defendants to further clarify both topics. Regarding the concept of testimonial injustice, this paper has three main implications. First, I pointed to a consequential ambiguity in the definition of testimonial injustice: it is unclear whether, to qualify for testimonial injustice, a negative identity prejudice must pertain to the testifier's capacity as a knower per se, or whether a negative identity prejudice can pertain to a different aspect of the testifier's identity. Considering the two canonical cases, I argued in favor of understanding testimonial injustice more broadly to include general prejudices. Second, I explained the parallels between Rivera's case and the canonical case from *To Kill a Mockingbird*. Through this comparison, I demonstrated that Rivera's case is an instance of credibility excess testimonial injustice, thereby supporting the inclusion of such a category in the definition. Third, I demonstrated that whether a person experiences a credibility deficit or excess corresponds to whether the content of the testimony confirms or undermines the pre-existing prejudice. As such, testifiers are similarly harmed and objectified, regardless of whether their testimony is adopted or ignored.

[196] Of the 375 people exonerated with DNA evidence in the past three decades in the US, 44 of them took plea deals instead of going to trial. See: "DNA Exonerations in the United States."

Regarding criminal defendants, I identified a previously unarticulated way in which criminal defendants may be harmed at trial. Through the refined theory of testimonial injustice, I argued that defendants may be especially vulnerable to testimonial injustice if jurors possess certain negative prejudicial beliefs. I also demonstrated how widespread silencing is a further epistemic harm experienced by defendants. Regardless of whether defendants are guilty or innocent, they may be epistemically harmed at trial. I also outlined potential non-epistemic harms that follow and demonstrated their wide-reaching effects.

Epistemic harms during and around criminal trials are especially concerning because trials have an explicitly epistemic purpose. Jurors are tasked with evaluating the evidence in front of them in order to determine guilt. If their own prejudices are leading them to improperly weigh testimony, then jurors are failing at their task. If testimonial injustice as a result of prejudices against criminal defendants is indeed common, it would seriously undermine the presumption of innocence supposedly afforded to criminal defendants, and perhaps, more broadly, undermine the legitimacy of a punishment system that relies on jury trials.

Bibliography

Blume, John H. "The Dilemma of the Criminal Defendant with a Prior Record-Lessons from the Wrongfully Convicted." *Journal of Empirical Legal Studies* 5, no. 3 (2008): 477–505. https://doi.org/10.1111/j.1740-1461.2008.00131.x.

Burnett, Ros, Carolyn Hoyle, and Naomi-Ellen Speechley. "The Context and Impact of Being Wrongly Accused of Abuse in Occupations of Trust." *The Howard Journal of Crime and Justice* 56, no. 2 (2017): 176–97. https://doi.org/10.1111/hojo.12199.

Costanzo, Mark, Netta Shaked-Schroer, and Katherine Vinson. "Juror Beliefs About Police Interrogations, False Confessions, and Expert Testimony." *Journal of Empirical Legal Studies* 7, no. 2 (2010): 231–47. https://doi.org/10.1111/j.1740-1461.2010.01177.x.

Davis, Emmalon. "Typecasts, Tokens, and Spokespersons: A Case for Credibility Excess as Testimonial Injustice." *Hypatia* 31, no. 3 (2016): 485–501. https://doi.org/10.1111/hypa.12251.

Dembroff, Robin, and Dennis Whitcomb. "Content Focused Epistemic Injustice." In *Oxford Studies in Epistemology*, edited by Tamar Szabó Gendler, John Hawthorne, and Julianne Chung, 7:48–70. Oxford University Press, 2022. https://doi.org/10.1093/oso/9780192868978.003.0003.

Demuth, Stephen, and Darrell Steffensmeier. "Ethnicity Effects on Sentence Outcomes in Large Urban Courts: Comparisons Among White, Black, and Hispanic Defendants." *Social Science Quarterly* 85, no. 4 (2004): 994–1011. https://doi.org/10.1111/j.0038-4941.2004.00255.x.

Espinoza, Russ K. E., and Cynthia Willis-Esqueda. "Defendant and Defense Attorney Characteristics and Their Effects on Juror Decision Making and Prejudice against Mexican Americans." *Cultural Diversity and Ethnic Minority Psychology* 14, no. 4 (2008): 364–71. https://doi.org/10.1037/a0012767.

Fricker, Miranda. *Epistemic Injustice: Power & the Ethics of Knowing. Epistemic Injustice*. Oxford Academic, 2007. https://doi.org/10.1093/acprof:oso/9780198237907.001.0001.

Friedman, H.M. "Destigmatizing the Innocent Defendant: Rehabilitating Reputation After an Assault Under Color of Law." *Criminal Law Bulletin* 28, no. 2 (1992): 116–36.

Geffen, Benjamin D. "The Collateral Consequences of Acquittal: Employment Discrimination on the Basis of Arrests without Convictions." *University of Pennsylvania Journal of Law and Social Change* 20, no. 2 (2017): 81–104.

Gonzalez, Eduardo. "Stereotypical Depictions of Latino Criminality: U.S. Latinos in the Media during the MAGA Campaign." *Democratic Communiqué* 28, no. 1 (2019): 46–62.

Gramlich, John. "Only 2% of Federal Criminal Defendants Go to Trial, and Most Who Do Are Found Guilty." *Pew Research Center* (blog). Accessed November 10, 2021. https://www.pewresearch.org/fact-tank/2019/06/11/only-2-of-federal-criminal-defendants-go-to-trial-and-most-who-do-are-found-guilty/.

Hornsby, Jennifer, and Rae Langton. "Free Speech and Illocution." *Legal Theory* 4, no. 1 (March 1998): 21–37. https://doi.org/10.1017/S1352325200000902.

Innocence Project. "DNA Exonerations in the United States." Accessed November 11, 2021. https://innocenceproject.org/dna-exonerations-in-the-united-states/.

Kassin, S. M., and K. Neumann. "On the Power of Confession Evidence: An Experimental Test of the Fundamental Difference Hypothesis." *Law and Human Behavior* 21, no. 5 (October 1997): 469–84. https://doi.org/10.1023/a:1024871622490.

Krishnamurthi, Guha. "The Case for the Abolition of Criminal Confessions." *SMU Law Review* 75, no. 1 (2022): 15–72. https://doi.org/10.2139/ssrn.3730499.

Lackey, Jennifer. *Criminal Testimonial Injustice*. Oxford: Oxford University Press, 2023. https://doi.org/10.1093/oso/9780192864109.001.0001.

———. "False Confessions and Agential Testimonial Injustice." In *Criminal Testimonial Injustice*, 31–72. Oxford University Press, 2023. https://doi.org/10.1093/oso/9780192864109.003.0003.

———. "False Confessions and Subverted Agency." *Royal Institute of Philosophy Supplement* 89 (2021): 11–35. https://doi.org/10.1017/S1358246121000072.

———. "False Confessions and Testimonial Injustice." *Journal of Criminal Law and Criminology* 110, no. 1 (2020): 43–68.

Lee, Harper. *To Kill a Mockingbird*. London: William Heinemann, 1960.

Leo, Richard, and Steven Drizin. "The Problem of False Confessions in the Post-DNA World." *North Carolina Law Review* 82 (2008).

Loevy, Debra. "Juan Rivera's Coerced False Confession." *Loevy & Loevy* (blog), May 30, 2015. https://www.loevy.com/blog/juan-riveras-coerced-false-confession/.

Luzzi, Federico. "Testimonial Injustice Without Credibility Deficit (or Excess)." *Thought: A Journal of Philosophy* 5, no. 3 (2016): 203–11. https://doi.org/10.1002/tht3.212.

Mayson, Sandra. "Dangerous Defendants." *Yale Law Journal* 127, no. 3 (2017): 490–569.

McGlynn, Aidan. "Epistemic Objectification as the Primary Harm of Testimonial Injustice." *Episteme* 18, no. 2 (2021): 160–76. https://doi.org/10.1017/epi.2019.9.

Medina, José. "The Relevance of Credibility Excess in a Proportional View of Epistemic Injustice: Differential Epistemic Authority and the Social Imaginary." *Social Epistemology* 25, no. 1 (2011): 15–35. https://doi.org/10.1080/02691728.2010.534568.

Minghella, Anthony. *The Talented Mr. Ripley: A Screenplay*. London: Bloomsbury Methuen Drama, 2014.

Natapoff, Alexandra. "Speechless: The Silencing of Criminal Defendants." *New York University Law Review* 80, no. 5 (2005): 1449–1504.

O'Donnell, Christina, and Martin Safer. "Jury Instructions and Mock-Juror Sensitivity to Confession Evidence in a Simulated Criminal Case." *Psychology, Crime & Law* 23 (2017): 1–47. https://doi.org/10.1080/1068316X.2017.1351965.

Pohlhaus, Gaile. "Discerning the Primary Epistemic Harm in Cases of Testimonial Injustice." *Social Epistemology* 28, no. 2 (2014): 99–114. https://doi.org/10.1080/02691728.2013.782581.

Schulhofer, Stephen J. "Some Kind Words for the Privilege against Self-Incrimination the Bill of Rights Yesterday and Today." *Valparaiso University Law Review* 26, no. 1 (1991): 311–36.

Strub, Tanya, and Blake M. McKimmie. "Sugar and Spice and All Things Nice: The Role of Gender Stereotypes in Jurors' Perceptions of Criminal Defendants." *Psychiatry, Psychology and Law* 23, no. 4 (2016): 487–98. https://doi.org/10.1080/13218719.2015.1080151.

Subramanian, Ram, Léon Digard, Melvin Washington, and Stephanie Sorage. "In the Shadows: A Review of the Research on Plea Bargaining." Vera Institute, September 2020.

Wanderer, Jeremy. "Addressing Testimonial Injustice: Being Ignored and Being Rejected." *The Philosophical Quarterly* 62, no. 246 (2012): 148–69. https://doi.org/10.1111/j.1467-9213.2011.712.x.

Yap, Audrey S. "Credibility Excess and the Social Imaginary in Cases of Sexual Assault." *Feminist Philosophy Quarterly* 3, no. 4 (2017). https://doi.org/10.5206/fpq/2017.4.1.

Statistical Evidence and the Criminal Verdict Asymmetry

Epistemologists have posed the following puzzle, known as the proof paradox: Why is it intuitively problematic for juries to convict on the basis of statistical evidence and yet intuitively unproblematic for juries to convict on the basis of far less reliable, non-statistical evidence? To answer this question, theorists have explained the exclusion of statistical evidence by arguing that legal proof requires certain epistemic features. In this paper, I make two contributions to the debate. First, I establish the Criminal Verdict Asymmetry, a previously unarticulated asymmetry between the epistemic norms of guilty and not guilty verdicts. I argue that the prosecution and defense's different epistemic burdens influence whether statistical evidence can generate the type of verdict each side pursues. Second, I point out a mistake in how theorists have understood the role of statistical evidence in criminal trials. Though epistemologists have primarily focused on whether statistical evidence can generate specific epistemic features required for convictions, I consider whether statistical evidence can demonstrate a lack of such features. I find that there are epistemic advantages to allowing the defense to introduce statistical evidence that reveals the prosecution's failure to prove the defendant's guilt.

Introduction

Otis Boone, a young Black man, was accused of committing two robberies in the United States.[197] Both robbery victims were White. In the police lineup, the police brought in Boone along with a few other people who matched the physical description provided by the victims. Both victims identified Boone as the person who had stolen their phones, and there was no physical evidence in the case. He was convicted and sentenced to 25 years in prison. On appeal, Boone was granted a retrial in which he successfully demonstrated to the jury that he had been many miles away from both crime locations at the times when each respective robbery took place. After serving seven years in prison, he was finally exonerated.[198]

Eyewitness misidentification landed Boone in prison for crimes he did not commit. The mistakes in Boone's case are potential examples of the "cross-race effect." The cross-race effect is the

[197] Southall, "A Black Man Spent 7 Years in Prison. Then a Court Changed the Rules on Racial Bias."
[198] All future mentions of this case are also from Southall.

psychological phenomenon that people often misidentify people of other races.[199] In fact, approximately 70% of those exonerated through DNA evidence in the United States were originally convicted due to witnesses who identified the wrong assailant.[200]

But whether or not juries may be warned about the cross-race effect when evaluating evidence during trial is debated.[201] Indeed, in Boone's case, the court prohibited the jury from hearing about the cross-race effect since it constituted introducing "bare statistical evidence," also known as "naked statistical evidence."[202]

Though a few scholars support the inclusion of statistical evidence across the board, the more commonly-held view is that there is something troubling about statistical evidence in legal contexts, and especially so in the context of criminal trials.[203] The concerns regarding statistical evidence in legal contexts fall roughly into two categories: epistemic and moral. There is a further split between the

---

[199] Connelly, "Cross-Racial Identifications: Solutions to the 'They All Look Alike' Effect."

[200] Agathocleus, "How Eyewitness Misidentification Can Send Innocent People to Prison."

[201] For an overview of the inadmissibility of such evidence in different legal jurisdictions in the US, see Fradella, "Why Judges Should Admit Expert Testimony on the Unreliability of Eyewitness Testimony." For studies about the cross-race effect in court, see, e.g., Connelly, "Cross-Racial Identifications: Solutions to the 'They All Look Alike' Effect"; Johnson, "Cross-Racial Identification Errors in Criminal Cases"; Wilson, Hugenberg, and Bernstein, "The Cross-Race Effect and Eyewitness Identification"; Abshire and Bornstein, "Juror Sensitivity to the Cross-Race Effect"; Woocher, "Did Your Eyes Deceive You--Expert Psychological Testimony on the Unreliability of Eyewitness Identification Note"; Loftus, "Impact of Expert Psychological Testimony on the Unreliability of Eyewitness Identification."

[202] The precise definition statistical evidence is debated. Some scholars take a "I'll know it when I see it" approach, see, e.g., Enoch and Spectre, "Sensitivity, Safety, and the Law," 184. Roush (forthcoming) provided the following straightforward definition: "Naked statistic is a frequency probability for a population, a mere correlation between two properties."

[203] See, e.g., Allen, "On the Significance of Batting Averages and Strikeout Totals"; Blome-Tillmann, "Sensitivity, Causality, and Statistical Evidence in Courts of Law"; Blome-Tillmann, "'More Likely Than Not' - Knowledge First and the Role of Statistical Evidence in Courts of Law"; Blome-Tillmann, "Statistical Evidence, Normalcy, and the Gatecrasher Paradox"; Bolinger, "The Rational Impermissibility of Accepting (Some) Racial Generalizations"; Cheng, "Reconceptualizing the Burden of Proof"; Cohen, *The Probable and the Provable*; Dahlman, "Naked Statistical Evidence and Incentives for Lawful Conduct"; Dant, "Gambling on the Truth: The Use of Purely Statistical Evidence as a Basis for Civil Liability"; Di Bello, "Trial by Statistics"; Di Bello, "Proof Paradoxes and Normic Support"; Enoch, Spectre, and Fisher, "Statistical Evidence, Sensitivity, and the Legal Value of Knowledge"; Enoch and Fisher, "Sense and Sensitivity"; Fienberg and Schervish, "The Relevance of Bayesian Inference for the Presentation of Statistical Evidence and for Legal Decision Making"; Gardiner, "Legal Burdens of Proof and Statistical Evidence"; Gardiner, "Profiling and Proof"; Colyvan, Regan, and Ferson, "Is It a Crime to Belong to a Reference Class"; Koehler, "When Do Courts Think Base Rate Statistics Are Relevant?"; Moss, *Probabilistic Knowledge*; Moss, "Knowledge and Legal Proof"; Nance, *The Burdens of Proof*; Nesson, "Evidence or the Event - On Judicial Proof and the Acceptability of Verdicts"; Nunn, "The Incompatibility of Due Process and Naked Statistical Evidence"; Pardo, "The Field of Evidence and the Field of Knowledge"; Pardo, "The Gettier Problem and Legal Proof"; Pardo, "Safety vs. Sensitivity"; Pardo, "The Paradoxes of Legal Proof"; Pundik, "Statistical Evidence and Individual Litigants: A Reconsideration of Wasserman's Argument from Autonomy"; Pundik, "The Epistemology of Statistical Evidence"; Pundik, "Rethinking the Use of Statistical Evidence to Prove Causation in Criminal Cases"; Redmayne, "Exploring the Proof Paradoxes"; Smith, "When Does Evidence Suffice for Conviction?"; Thomson, "Liability and Individualized Evidence"; Tribe, "Trial by Mathematics"; Twining, "Debating Probabilities"; Wasserman, "The Morality of Statistical Proof and the Risk of Mistaken Liability Decision and Interference Litigation"; Wright, "Causation, Responsibility, Risk, Probability, Naked Statistics, and Proof." For a summary, see: Ross, "Recent Work on the Proof Paradox."

types of arguments presented regarding statistical evidence: philosophers tend to discuss the sufficiency of statistical evidence for conviction, while legal scholars tend to discuss the admissibility of statistical evidence in court.[204] This paper primarily considers *epistemic* arguments regarding the *sufficiency* of statistical evidence in criminal trials.

In this paper, I make two connected contributions to the debate on the exclusion of statistical evidence in criminal cases. First, I establish a previously unarticulated asymmetry between the epistemic norms of guilty and not guilty verdicts. This asymmetry, which I term the **Criminal Verdict Asymmetry**, reveals that the defense and prosecution have different aims—while the prosecution aims to demonstrate that the defendant is guilty, the defense only aims to demonstrate that the prosecution has failed to prove that the defendant is guilty. The defense and prosecution, therefore, have different epistemic burdens, which could influence whether statistical evidence can generate the type of verdict each side pursues.

Second, I argue that statistical evidence introduced by the *defense* can serve an important role in criminal trials. Epistemologists have primarily focused on whether statistical evidence can generate specific epistemic features required for convictions. There is widespread consensus that statistical evidence alone is insufficient for legal proof. Theorists, however, have not yet considered whether statistical evidence can helpfully demonstrate a *lack* of legal proof. This type of statistical evidence, given its role, can be understood as "debunking statistical evidence."[205] Given the **Criminal Verdict Asymmetry**, I argue that defense-introduced debunking statistical evidence can be sufficient for an acquittal.

In §1, I lay out the legal epistemological puzzle of statistical evidence, known as the proof paradox. In §2, I develop and explain the **Criminal Verdict Asymmetry**. In §3, I present a Gettier case to demonstrate how statistical evidence can reveal that judgments regarding certain propositions are unreliable. In §4, I explain how statistical evidence in criminal cases could assist the defense in showing that the prosecution has failed to prove the defendant's guilt. In §5, I raise and address two concerns.

[204] For a more robust discussion of these two types of distinctions, see Dahlman and Pundik, "The Problem with Naked Statistical Evidence."
[205] I thank one of the anonymous reviewers at Synthese for suggesting the term "debunking."

§1. Proof Paradox

First, let us establish the puzzle of statistical evidence in legal contexts, known as the proof paradox. Consider the following case, based on an example first introduced by Charles Nesson:[206]

> **Prison Yard**: 100 people are exercising in a prison yard, when 99 of them suddenly join together in a planned attack on the prison guards. The remaining person tries to stop the attack. There is no available evidence distinguishing the innocent person from the rest. Local prosecutors randomly select one person, Polly, and bring her to trial for participating in the attack.

In this case, there is no way to distinguish between those who rioted and the one person who did not participate. There is a 99% chance that Polly participated in the riot. Many people, however, share the strong intuition that it would be problematic to convict Polly solely on the basis of that statistical evidence. The exclusion of such statistical evidence in criminal cases is especially puzzling because legal systems permit other far less reliable forms of evidence, like eyewitness accounts.

Why is it intuitively problematic for juries to convict on the basis of statistical evidence? To answer this question, theorists have explained the exclusion of statistical evidence by arguing that legal proof requires certain epistemic features, such as sensitivity,[207] safety,[208] normic support,[209] or knowledge requirements.[210]

A first possible solution to the paradox is that legal proof requires sensitivity, namely that if P were false, then we would not believe P.[211] In Polly's case, the belief that Polly is guilty does not meet the sensitivity requirement. If Polly was in fact innocent, we would still believe that Polly was guilty based on the evidence. A second possible explanation is that legal proof requires safety, namely that P is also true in nearby worlds.[212] In Polly's case, the belief that Polly is guilty is not safe since there is a nearby world—namely the world in which Polly was the one person who did not participate—in which P is false. A third explanation is that evidence needs to provide normic support for legal

---

[206] This case is adapted from Nesson's well-known case. See Nesson, "Reasonable Doubt and Permissive Inferences," 1192–94.

[207] Enoch, Spectre, and Fisher, "Statistical Evidence, Sensitivity, and the Legal Value of Knowledge"; Enoch and Fisher, "Sense and Sensitivity."

[208] Pardo, "Safety vs. Sensitivity"; Pritchard, "Anti-Luck Epistemology and the Gettier Problem." For the purpose of this paper, I use the term "safety" in a broad sense, as I remain agnostic regarding the precise definition of safety which is debated extensively.

[209] Smith, "When Does Evidence Suffice for Conviction?"

[210] Moss, "Knowledge and Legal Proof."

[211] Enoch, Spectre, and Fisher, "Statistical Evidence, Sensitivity, and the Legal Value of Knowledge"; Enoch and Fisher, "Sense and Sensitivity."

[212] Pardo, "Safety vs. Sensitivity"; Pritchard, "Anti-Luck Epistemology and the Gettier Problem."

proof.[213] A piece of evidence provides normic support for a proposition if a further explanation would be required if the proposition turned out to be false. Statistical evidence does not provide normic support for the belief that Polly is guilty, since if the belief were to be false and Polly was in fact innocent, no further explanation would be required.

Sarah Moss provides another possible solution. She argues that legal proof requires knowledge.[214] If legal proof requires knowledge, the exclusion of statistical evidence can be explained since statistical evidence cannot generate knowledge.[215] To demonstrate why the statistical evidence alone cannot generate legal proof in the **Prison Yard** case, Moss points out parallels between Nesson's case and a well-known case originally presented by Gilbert Harman:[216]

> **Lottery**: Lottie has purchased a ticket in a small lottery. There are 100 tickets in all. The winning number has already been selected at random, but it hasn't been announced yet.

In this case, there is no way to know whether Lottie has a winning or a losing ticket. There is a 99% chance that Lottie lost the lottery. But if Lottie cannot rule out the possibility that she has the one winning ticket, Lottie cannot *know* that she lost. In both **Prison Yard** and **Lottery**, one may form a likely true belief on the basis of the available statistical evidence. Yet, as long as there is a live possibility that Polly is innocent or that Lottie's ticket is the winner, one cannot know that Polly participated in the riot or that Lottie lost the lottery.

Moss also contends that when a jury finds a person guilty, the jury asserts their finding by declaring the person guilty. The speech act is an assertion. It would be confusing for the jury to say: "The defendant might be innocent. We find him guilty." Moss argues that the issue with this assertion is that knowledge is the norm of assertion.[217] Thus, if the jury does not know that the defendant is guilty, then the jury should not assert that the defendant is guilty.

Though this list of proof paradox solutions is not meant to be exhaustive, all of these solutions point to certain epistemic features that may be required for conviction. For simplicity purposes, I will adopt the knowledge solution.[218] Along the way, however, I will point out how the other solutions

---

[213] Smith, "When Does Evidence Suffice for Conviction?"
[214] Moss, "Knowledge and Legal Proof."
[215] Statistical evidence is unable to generate first-order knowledge, though it may be able to yield higher-order knowledge which I discuss below.
[216] This case is adapted from Harman's well-known case that established what is known as the lottery paradox. See Harman, *Change in View: Principles of Reasoning*, 71.
[217] Moss, "Knowledge and Legal Proof."
[218] The knowledge solution is more stringent than the other solutions since knowledge is factive. Moreover, the knowledge solution is distinct from the others since it explains what is both necessary and sufficient for verdicts, while the other views only explain what type of evidence is necessary for verdicts.

similarly apply. In the coming section, I will discuss the epistemic difference between what the prosecution and defense aim to prove.

§2. Criminal Verdict Asymmetry

There is an important, yet previously unarticulated, asymmetry between the epistemic norms of guilty and not guilty verdicts, which I term the **Criminal Verdict Asymmetry**. To demonstrate the asymmetry, let us first begin with the aims of the prosecution. The prosecution aims to prove that the defendant is guilty. Since the prosecution carries the burden of proof, they need to present enough evidence to the jury so that the jury is in a position to find the defendant guilty.

At first glance, it may appear that the defense aims to prove that the defendant is innocent. When a jury finds a person guilty, the jury declares: "We find the defendant guilty." A guilty and a not guilty verdict are structurally the same. The two verdicts are parallel statements, one the affirmative and one the negative. If a guilty verdict is a declaration of guilt, similarly one may conceive of a not guilty verdict as a declaration of innocence. Indeed, if the defendant is found not guilty, the jury declares: "We find the defendant not guilty." This formulation of the verdict makes it seem that the jury is declaring that the defendant is not guilty.

I argue, however, that the defense is not trying to prove that the defendant is innocent. Instead, the defense only needs to demonstrate that the prosecution has failed to prove that the defendant is guilty. The jury declaration of a not guilty verdict is best understood as follows: "(Not) We find the defendant guilty." Though the placement of the "not" in the typical verdict declaration can be misleading, it is important to recognize that the defense does not aim to prove that the defendant is innocent. While the prosecution aims to prove guilt, the defense only aims to demonstrate that the prosecution has failed to meet its burden. Nothing additional is necessary for the defense to accomplish its goal of acquitting the defendant.

While this difference appears fairly straightforward, it means that the prosecution and the defense have different epistemic burdens. The difference creates an asymmetry between the epistemic status of guilty and not guilty verdicts. On the one hand, the prosecution aims to prove the defendant's guilt. On the other hand, the defense only aims to demonstrate that the prosecution has failed to prove the defendant's guilt. This asymmetry impacts the type of evidence that can be used to produce each outcome. If the prosecution aims to prove the defendant is guilty, then statistical evidence is insufficient for this aim as demonstrated by the proof paradox. If the defense, however, is only trying to demonstrate that the prosecution has failed to prove the defendant's guilt, perhaps statistical

evidence could be introduced to yield a not guilty verdict. In the coming sections, I explore this possibility.

§3. Statistical Evidence in Gettier Cases

In discussions about the use of statistical evidence in criminal cases, scholars typically focus on why statistical evidence is insufficient for criminal convictions. I argue, however, that statistical evidence can reveal the unreliability of judgments regarding certain propositions which can be a helpful tool for the defense.

To demonstrate how statistical evidence operates relative to different propositions, let us consider the relationship between statistical evidence and potential beliefs in a non-legal case. Consider this classic Gettier-style case:[219]

> **Barnsville**: As Brenda is driving through a town, she sees a barn. As it happens, the structure she sees is indeed a barn. However, suppose that Brenda's friend, Cara, later tells her that Barnsville is known for its fake barns, since 90% of the structures that appear to be barns are in fact fake barns. The fake barns look identical to real barns from the road, but they are actually just facades.[220] Furthermore, Cara is a reliable and accurate source.

In this case, Brenda originally believed P, namely that the barn she saw was real. Cara then provided a piece of statistical evidence Q, namely that 90% of the structures that appear to be barns in that town are in fact fake barns. Once Brenda hears Q, Brenda no longer believes P and she may even believe ~P, namely that the barn that she saw was fake.

Q is statistical evidence for ~P. Q means that it is highly likely that the barn that Brenda saw was fake. However, one cannot know that the barn that Brenda saw was fake merely on the basis of Q. For the same reason that Lottie cannot know that she lost the lottery and that the prosecutors cannot know that Polly participated in the riot, Brenda similarly cannot know on the basis of Q that the barn she saw was fake. Indeed, in **Barnsville**, the barn that Brenda saw was actually a real barn. Thus, ~P is not even true. But even if Brenda had seen a fake barn, one could not know ~P on the basis of Q alone since statistical evidence by itself cannot generate knowledge.[221]

---

[219] The fake barn case was first introduced in Ginet, "The Fourth Condition", as an example of the type of case presented by Gettier. See Gettier, "Is Justified True Belief Knowledge?"

[220] There is nothing wrong with Brenda's perception since the real and fake barns look identical. Any person in Barnsville would have been unable to distinguish between the two types of barns.

[221] This is true of first-order knowledge, for instance knowledge of P.

Importantly, Q provides direct evidence for a different proposition: Brenda does not know that P. Direct evidence can be understood as evidence for a proposition when the evidence entails that the proposition is true. Thus, Q is direct evidence for the proposition that Brenda does not know that P. Q indicates that Brenda's belief in P is not a safe belief since it reveals that P is not true in a close-by possible world. If Brenda had looked at any given barn-looking structure in town, Brenda likely would not have seen a real barn. Thus, Q reveals that Brenda's original belief in P is not a safe belief. If Brenda's belief in P is not safe, then Brenda cannot know P if safety is a necessary condition for knowledge.[222] Q therefore entails that Brenda does not know that P. Thus, on the basis of Q, we know that Brenda does not know that P.[223]

Though Q is statistical evidence, it is crucial to consider its relation to a given proposition. In relation to the proposition that Brenda saw a fake barn, Q does not generate knowledge. However, in relation to the proposition that Brenda does not know she saw a real barn, Q serves as direct evidence since Q reveals that Brenda's belief in P is not a safe belief. We therefore know that Brenda does not know that she saw a real barn. While relative to one proposition Q cannot generate knowledge, Q does generate knowledge relative to a different proposition.

Let us quickly consider how some of the other proof paradox solutions apply in this case. As discussed above, Q reveals that Brenda's belief in P is not safe. Similarly, Q also reveals that Brenda's belief in P is not sensitive. If Brenda had seen a fake barn, she would have originally still believed P, namely that she saw a real barn. Thus, even if P was false, Brenda would still believe that P. Finally, Q does not provide normic support for P.[224] With Q in mind, one would not require any further explanation if P were false since it would be unsurprising that Brenda saw one of the fake barns. Now that we have established how statistical evidence works in a Gettier case, let us turn to the use of statistical evidence in criminal cases.

---

[222] For the purpose of this paper, I adopt the widely held view that safety is a necessary condition for knowledge. A version of my argument, however, would similarly apply to the other views, such as those who reject luck-based knowledge. See, e.g., Pritchard, "Anti-Luck Epistemology and the Gettier Problem.".

[223] In this instance, statistical evidence appears to yield higher-order knowledge about what we can and cannot know in this case.

[224] Smith's notion of normic support addresses the relationship between a piece of evidence and a proposition—not the epistemic status of the proposition itself. The normic support view does not take a stance on the status of Brenda's belief that she saw a real barn. Rather, it can only be used to evaluate whether Q supports P.

§4. Statistical Evidence in Criminal Cases

Let us return to the Boone case. To review, Boone was convicted of two robberies that he did not commit. He was identified as the assailant by both victims in a police lineup. There was no physical evidence in the case so the eyewitness identification by the victims was the only evidence that connected Boone to the robberies.

Boone and the actual assailant were of a different race than that of the witnesses who mistakenly identified Boone. Possibly because of the cross-race effect, the witnesses mistook Boone for their actual assailant. The eyewitness testimony served as evidence for the proposition that Boone was guilty of committing the two robberies. Let us clarify the relevant propositions so that we can understand the evidence for each proposition.

A: Boone is guilty.

B: The eyewitnesses correctly identified Boone as the assailant.

In this case, the prosecution aimed to prove A. B, namely the proposition that both victims correctly identified Boone as the assailant, was strong evidence for A. If the jury believed B, it is understandable that this would lead them to believe A. Furthermore, since there was no other evidence in the case, B served as the primary, if not only, basis for believing A.

The defense was not permitted to introduce evidence regarding the cross-race effect because it was considered a form of statistical evidence. Though psychologists have not conclusively studied how frequently White witnesses misidentify Black suspects, let us suppose that an application of the cross-race effect could be translated into the following piece of statistical evidence:

C: 90% of White witness misidentify Black suspects.

In Boone's case, the eyewitnesses were White and Boone, the suspect, was Black. If C had been introduced, it would have provided support for ~B, namely that the eyewitness did not correctly identify Boone as the assailant. Indeed, ~B is true. Moreover, if the jury believed ~B, then there would not have been evidence for A since the eyewitness identification was the only piece of evidence presented in the case. If the jury had no evidence to support A, then the jury would have likely found Boone not guilty.

C, however, is merely statistical evidence for ~B. C means that it is likely that the White eyewitnesses misidentified Boone. However, one cannot know for certain whether Boone was misidentified on the basis of C alone. Again, statistical evidence cannot produce knowledge.[225] As we

[225] The distinction between first-order and higher-order knowledge applies here as well.

saw in the **Barnsville** case, the fact that 90% of the barns in Barnsville were fake did not mean that Brenda knew that the barn was fake, and in fact, the barn that Brenda happened to see was real. C only provides statistical evidence for ~B, so one cannot know ~B on the basis of C alone. On the basis of similar reasoning, courts have commonly excluded pieces of statistical evidence like C, since it cannot generate knowledge of the propositions the evidence appears to support.[226]

If the defense were to introduce C, however, it would not need to prove ~B. It is not necessary for the defense to demonstrate that the White eyewitnesses did indeed misidentify Boone. If the defense aimed to prove that the defendant is innocent, then one may be concerned that the jury would mistakenly think they know ~B on the basis of C. As discussed earlier, the jury cannot know ~B on the basis of C alone. Per the **Criminal Verdict Asymmetry**, however, the defense does not need the jury to know that the defendant is innocent. The defense only needs to demonstrate that the jury does not know that the defendant is guilty. Thus, C is only necessary to demonstrate a lack of knowledge. In this case, C can demonstrate that the jury does not know B.

Indeed, C is direct evidence for the proposition that the jury does not know that B. C reveals that the jury's belief in B is not safe. In the majority of cases in which White witnesses identify Black suspects, the witnesses are mistaken. Thus, the jury cannot know that B, as the jury's belief in B is not safe. If safety is a necessary condition for knowledge, then C reveals that the jury does not know that B. Thus, C actually entails that the jury does not know that B. On the basis of C, we know that the jury does know that B.

Since B is the primary support for the jury's belief in A, namely that Boone is guilty, demonstrating that the jury does not know B influences A as well. If the jury does not know B since their belief in B is unsafe, then the jury does not know that A since their belief in A is exclusively reliant on an unsafe belief. If the jury does not know that A, then the jury cannot and should not find Boone guilty.

Indeed, it is not the defense's goal to prove that Boone is innocent. Rather, in order to obtain a not guilty verdict, the defense only needs to prove that the jury is not in a position to know, on the basis of the evidence presented by the prosecution, that Boone is guilty. Statistical evidence, like C, can demonstrate that the jury's belief in Boone's guilt is based on unsafe evidence. If so, then the jury cannot know that the defendant is guilty and therefore should declare Boone not guilty.

---

[226] See, e.g., Johnson, "Cross-Racial Identification Errors in Criminal Cases"; Abshire and Bornstein, "Juror Sensitivity to the Cross-Race Effect"; Wilson, Hugenberg, and Bernstein, "The Cross-Race Effect and Eyewitness Identification."

The other solutions to the proof paradox generate a similar picture. As discussed, C reveals that the jury's belief in B is not safe. Similarly, C reveals that the jury's belief in B is not sensitive since if B was false, the jury would still believe that B. Indeed, in Boone's case, B was believed by the jury even though B was actually false since the witnesses misidentified Boone. Thus, the statistical evidence would have helped the jury realize that their belief in B is not sensitive. Similarly, C does not normically support B. Given C, no further explanation is required to understand why B is false. Simply put, it is unsurprising that the witnesses misidentified Boone given the cross-race effect. In contrast, consider a case in which the witnesses, the suspect, and the defendant were all of the same race. If the witnesses misidentified the suspect and the defendant was actually innocent, one would expect an explanation. In Boone's case, however, no further explanation is required since the cross-race effect explains the misidentification.

Importantly, debunking statistical evidence like C can be *sufficient* for a not guilty verdict. If the defense at Boone's trial had been permitted to introduce debunking statistical evidence like C, then the jury would have realized that their belief in Boone's guilt on the basis of B was insufficient for conviction. Even without additional evidence introduced by the defense, the debunking statistical evidence alone would have been sufficient to reach a not guilty verdict.

The **Criminal Verdict Asymmetry** reveals a difference in the types of evidence that can assist each side. Since the prosecution is trying to prove the defendant's guilt, the prosecution cannot use statistical evidence as demonstrated by the proof paradox. The defense, however, is merely trying to demonstrate that the prosecution failed to prove the defendant's guilt. Statistical evidence can demonstrate that certain pieces of evidence introduced by the prosecution are unreliable. If the defense can successfully demonstrate that the jury's belief is based on unreliable evidence, then the defense has accomplished its goal. Thus, while the prosecution cannot exclusively use statistical evidence to reach a conviction, the defense could exclusively use debunking statistical evidence to reach an acquittal.

§5. Sufficiency & Admissibility

Thus far, I have argued that there are epistemic benefits to defense-introduced statistical evidence. Though statistical evidence is not sufficient for conviction, debunking statistical evidence could be sufficient for acquittal.

At the beginning of the paper, I put aside the questions about the admissibility of statistical evidence in criminal trials. Though this paper does not aim to mount a comprehensive argument in

favor of allowing the defense to introduce debunking statistical evidence, some of the arguments addressed here do lay the groundwork for such an argument.

One of the primary reasons why statistical evidence has been excluded from criminal trials is grounded in sufficiency concerns. As demonstrated by the **Prison Yard** case, statistical evidence on its own cannot yield a proper conviction. Even if statistical evidence is not the only piece of evidence introduced, there is another sufficiency concern: statistical evidence cannot properly take a body of evidence from one that is insufficient for conviction to a body of evidence that is sufficient for conviction.

My argument, however, reveals that statistical evidence can be sufficient for the defense's aim. As demonstrated by the analysis of Boone's case, the defense only needed to introduce debunking statistical evidence in order to yield an acquittal. Moreover, statistical evidence can transform a body of evidence that appears to be sufficient for conviction into a body of evidence that is clearly not sufficient for conviction. In Boone's case, for instance, the body of evidence, namely the eyewitness identifications, appeared to be sufficient for conviction. If the statistical evidence regarding the cross-race effect had been introduced, the statistical evidence would have revealed that the body of evidence was actually insufficient for conviction.

If the primary reason to exclude statistical evidence from criminal trials is due to sufficiency concerns, then the arguments here provide strong reasons to permit defense-introduced debunking statistical evidence. Moreover, as I demonstrated, statistical evidence can helpfully reveal the unreliability of certain pieces of evidence. Together, those two components provide a pro tanto epistemic reason for permitting such evidence in court. The drawbacks of permitting defense-introduced debunking statistical evidence—including moral and practical concerns—would need to be further explored.[227] I will briefly raise and address two relevant *admissibility* questions.

5.1 Uneven Acquittal Rates

Could allowing the defense to introduce debunking statistical evidence lead to uneven acquittal rates based on a person's race, ethnicity, nationality, religion, or membership in another group? For instance, suppose that there is a neighborhood that is predominantly Black and thus almost every eyewitness to a crime is also Black. In local trials, when the defendant is not Black and the eyewitness is Black, the defense team introduces statistical evidence regarding the cross-race effect, thereby increasing the

[227] Some scholars point to moral concerns that arise when using statistical evidence. See, e.g., Di Bello and O'Neil, "Profile Evidence, Fairness, and the Risks of Mistaken Convictions."

chance of acquittal of non-Black defendants. When the defendant is Black and the eyewitness is Black, however, the defense team does not introduce evidence of the cross-race effect since it is not relevant. As a result, non-Black defendants are acquitted at a higher rate in this neighborhood. Thus, one may be concerned that allowing debunking statistical evidence could lead to uneven acquittals rates on the basis of defendants' identities.

My response is three-fold. First, if the defense uses the cross-race effect as debunking statistical evidence in some cases but not others, it does not necessarily entail that there would be uneven acquittal rates. Rather, the debunking statistical evidence merely reveals that a certain piece of evidence presented by the prosecution is unreliable. The prosecution can and should introduce additional evidence if the prosecution aims to secure a conviction. Indeed, different amounts and types of evidence are required in different cases. This is true regardless of whether debunking statistical evidence may be introduced. For instance, if the prosecution has a video recording of the crime, there may be very little additional evidence required. In a different case, the prosecution may need to present multiple witnesses and forensic evidence in order to secure a conviction. Thus, the varied nature of the amount of additional evidence required is not a concern. Defense-introduced debunking statistical evidence does not automatically lead to an acquittal. Rather, it serves to check the reliability of the evidence presented.

Second, even if defense-introduced debunking statistical evidence leads to extra acquittals, those defendants would not have been properly convicted on the basis of the evidence presented by the prosecution. Indeed, in many legal jurisdictions if a person is factually guilty, yet there is insufficient evidence to convict, then the person is acquitted.

Third, it is worthwhile to note that aiming for an "even" acquittal rate obscures a problematic underlying picture. There is no reason to prefer the current status quo. For instance, Black Americans are wrongfully convicted at a rate which is seven times higher than White Americans.[228] If permitting defense-introduced statistical evidence led to Black Americans being acquitted at a higher rate than White Americans, perhaps the overall conviction rates by race would actually be more accurate. Thus, the notion of an "even" acquittal rate is something that may not even be desirable. Rather, "only proper convictions" should be the goal.

[228] Gross, Possley, and Stephens, "Race and Wrongful Convictions in the United States."

5.2 Prosecution-Introduced Debunking

Could the prosecution be permitted to introduce debunking statistical evidence to counter defense-introduced evidence? One may be concerned that the defense would introduce statistical evidence using reference classes that are most advantageous to them.[229] As a result, the statistical evidence may be true yet misleading. For instance, suppose that the eyewitness is a White woman and the defendant is a Black woman, so the defense introduces the cross-race effect. Suppose that it is also the case that women correctly identify other women at a fairly high rate, despite the cross-race effect. One may be concerned that the defense would introduce the cross-race effect but not introduce the statistical evidence that women properly identify other women at a fairly high rate. Perhaps, then, the prosecution should be permitted to introduce debunking statistical evidence to counter defense-introduced statistical evidence.

I have two responses. First, it is worthwhile to note that the debunking statistical evidence introduced by the defense would be governed by general rules of evidence. Thus, the defense would likely not be permitted to introduce misleading or irrelevant evidence, even if the evidence is technically true.

Second, it is certainly conceivable that the prosecution would be permitted to introduce debunking statistical evidence. This question, however, returns us to the debate regarding the sufficiency of statistical evidence. Since statistical evidence cannot move a body of evidence from insufficient for conviction to sufficient for conviction, prosecution-introduced debunking statistical evidence is likely less useful than defense-introduced debunking statistical evidence. That being said, the question of whether both sides or only the defense should be permitted to introduce debunking statistical evidence calls for further discussion.

Conclusion

In this paper, I first established a previously unarticulated asymmetry between the epistemic norms of guilty and not guilty verdicts. While the prosecution aims to prove that the defendant is guilty, the defense only aims to show that the prosecution failed to prove that the defendant is guilty. This asymmetry reveals that statistical evidence may be used differently by the opposing sides. I then argued that debunking statistical evidence can demonstrate that the jury's belief in the defendant's guilt is based on unreliable evidence. Thus, even though statistical evidence is insufficient for conviction,

[229] I thank the anonymous reviewer at Synthese for raising this concern.

defense-introduced debunking statistical evidence can be sufficient for acquittal. If the legitimacy of a legal system relies on ensuring that convictions are correct, justified, not luck-based, and responsive to proper evidence, then the arguments presented in this paper provide strong epistemic and moral reasons in favor of permitting defense-introduced statistical evidence.

Bibliography

Abshire, Jordan, and Brian H. Bornstein. "Juror Sensitivity to the Cross-Race Effect." *Law and Human Behavior* 27, no. 5 (2003): 471–80. https://doi.org/10.1023/A:1025481905861.

Agathocleus, Alexis. "How Eyewitness Misidentification Can Send Innocent People to Prison." Innocence Project, April 15, 2020. https://innocenceproject.org/how-eyewitness-misidentification-can-send-innocent-people-to-prison/.

Allen, Ronald J. "On the Significance of Batting Averages and Strikeout Totals: A Clarification of the Naked Statistical Evidence Debate, the Meaning of Evidence, and the Requirement of Proof beyond a Reasonable Doubt." *Tulane Law Review* 65 (1991): 1093–1110.

Blome-Tillmann, Michael. "'More Likely Than Not' - Knowledge First and the Role of Statistical Evidence in Courts of Law." In *Knowledge First - Approaches in Epistemology and Mind*, edited by Adam Carter, Emma Gordon, and Benjamin Jarvis, 278–92. Oxford University Press, 2017.

———. "Sensitivity, Causality, and Statistical Evidence in Courts of Law." *Thought: A Journal of Philosophy* 4, no. 2 (2015): 102–12. https://doi.org/10.1002/tht3.163.

———. "Statistical Evidence, Normalcy, and the Gatecrasher Paradox." *Mind* 129, no. 514 (2020): 563–78. https://doi.org/10.1093/mind/fzaa007.

Bolinger, Renée Jorgensen. "The Rational Impermissibility of Accepting (Some) Racial Generalizations." *Synthese* 197, no. 6 (June 1, 2020): 2415–31. https://doi.org/10.1007/s11229-018-1809-5.

Cheng, Edward K. "Reconceptualizing the Burden of Proof." *Yale Law Journal* 122 (2013): 1254–79.

Cohen, L. Jonathan. *The Probable and the Provable*. Clarendon Library of Logic and Philosophy. Oxford, New York: Oxford University Press, 1977.

Colyvan, Mark, Helen M. Regan, and Scott Ferson. "Is It a Crime to Belong to a Reference Class." *Journal of Political Philosophy* 9, no. 2 (2001): 168–81. https://doi.org/10.1111/1467-9760.00123.

Connelly, Laura. "Cross-Racial Identifications: Solutions to the 'They All Look Alike' Effect." *Michigan Journal of Race and Law* 21 (2015): 125–45.

Dahlman, Christian. "Naked Statistical Evidence and Incentives for Lawful Conduct." *The International Journal of Evidence & Proof* 24, no. 2 (2020): 162–79. https://doi.org/10.1177/1365712720913333.

Dahlman, Christian, and Amit Pundik. "The Problem with Naked Statistical Evidence." In *Philosophical Foundations of Evidence Law*, 2021. https://doi.org/10.2139/ssrn.3758860.

Dant, Mary. "Gambling on the Truth: The Use of Purely Statistical Evidence as a Basis for Civil Liability." *Columbia Journal of Law and Social Problems* 22, no. 1 (1988): 31–70.

Di Bello, Marcello. "Proof Paradoxes and Normic Support: Socializing or Relativizing?" *Mind* 129, no. 516 (2020): 1269–85. https://doi.org/10.1093/mind/fzz021.

———. "Trial by Statistics: Is a High Probability of Guilt Enough to Convict?" *Mind* 128, no. 512 (2019): 1045–84. https://doi.org/10.1093/mind/fzy026.

Di Bello, Marcello, and Collin O'Neil. "Profile Evidence, Fairness, and the Risks of Mistaken Convictions." *Ethics* 130, no. 2 (2020): 147–78. https://doi.org/10.1086/705764.

Enoch, David, and Talia Fisher. "Sense and Sensitivity: Epistemic and Instrumental Approaches to Statistical Evidence." *Stanford Law Review* 67 (2015): 557–611.

Enoch, David, and Levi Spectre. "Sensitivity, Safety, and the Law: A Reply to Pardo." *Legal Theory* 25, no. 3 (2019): 178–99. https://doi.org/10.1017/S1352325219000120.

Enoch, David, Levi Spectre, and Talia Fisher. "Statistical Evidence, Sensitivity, and the Legal Value of Knowledge." *Philosophy & Public Affairs* 40, no. 3 (2012): 197–224. https://doi.org/10.1111/papa.12000.

Fienberg, Stephen E., and Mark J. Schervish. "The Relevance of Bayesian Inference for the Presentation of Statistical Evidence and for Legal Decision Making." *Boston University Law Review* 66 (1986): 771–98.

Fradella, Henry F. "Why Judges Should Admit Expert Testimony on the Unreliability of Eyewitness Testimony." *Federal Courts Law Review* 2 (2007): 3–29.

Gardiner, Georgi. "Legal Burdens of Proof and Statistical Evidence." In *The Routledge Handbook of Applied Epistemology*, 179–95. Routledge, 2018.

———. "Profiling and Proof: Are Statistics Safe?" *Philosophy* 95, no. 2 (2020): 161–83. https://doi.org/10.1017/S0031819120000030.

Gettier, Edmund L. "Is Justified True Belief Knowledge?" *Analysis* 23, no. 6 (1963): 121–23. https://doi.org/10.1093/analys/23.6.121.

Ginet, Carl. "The Fourth Condition." In *Philosophical Analysis: A Defense by Example*, edited by David F. Austin and Edmund L. Gettier, 39:105–17. Dordrecht; London: Kluwer, 1988.

Gross, Samuel R, Maurice Possley, and Klara Stephens. "Race and Wrongful Convictions in the United States." University of California Irvine: National Registry of Exonerations, March 2017.

Harman, Gilbert. *Change in View: Principles of Reasoning*. Cambridge: MIT Press, 1986.

Johnson, Sheri Lynn. "Cross-Racial Identification Errors in Criminal Cases." *Cornell Law Review* 69, no. 5 (84 1983): 934–87.

Koehler, Jonathan J. "When Do Courts Think Base Rate Statistics Are Relevant?" *Jurimetrics* 42, no. 4 (2002): 373–402.

Loftus, Elizabeth F. "Impact of Expert Psychological Testimony on the Unreliability of Eyewitness Identification." *Journal of Applied Psychology* 65, no. 1 (1980): 9–15. https://doi.org/10.1037/0021-9010.65.1.9.

Moss, Sarah. "Knowledge and Legal Proof." In *Oxford Studies in Epistemology 7*, Vol. 7. Oxford University Press, 2021.

———. *Probabilistic Knowledge*. Oxford, New York: Oxford University Press, 2018.

Nance, Dale A. *The Burdens of Proof*. Cambridge University Press, 2016.

Nesson, Charles R. "Evidence or the Event - On Judicial Proof and the Acceptability of Verdicts." *Harvard Law Review* 98, no. 7 (1985): 1357–92.

———. "Reasonable Doubt and Permissive Inferences: The Value of Complexity." *Harvard Law Review* 92, no. 6 (1979): 1187–1225. https://doi.org/10.2307/1340444.

Nunn, G Alexander. "The Incompatibility of Due Process and Naked Statistical Evidence." *Vanderbilt Law Review* 68 (2015): 1406–33.

Pardo, Michael. "Safety vs. Sensitivity: Possible Worlds and the Law of Evidence." *Legal Theory* 24, no. 1 (2018): 50–75. https://doi.org/10.1017/S1352325218000010.

———. "The Field of Evidence and the Field of Knowledge." *Law and Philosophy* 24 (2005): 321–92. https://doi.org/10.1007/s10982-004-4999-6.

———. "The Gettier Problem and Legal Proof." *Legal Theory* 16, no. 1 (2010): 37–57. https://doi.org/10.1017/S1352325210000054.

———. "The Paradoxes of Legal Proof: A Critical Guide." *Boston University Law Review* 99, no. 1 (2019): 233–90.
Pritchard, Duncan. "Anti-Luck Epistemology and the Gettier Problem." *Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition* 172, no. 1 (2015): 93–111.
Pundik, Amit. "Rethinking the Use of Statistical Evidence to Prove Causation in Criminal Cases: A Tale of (Im)Probability and Free Will." *Law and Philosophy* 40, no. 2 (April 1, 2021): 97–128. https://doi.org/10.1007/s10982-020-09389-0.
———. "Statistical Evidence and Individual Litigants: A Reconsideration of Wasserman's Argument from Autonomy." *The International Journal of Evidence & Proof* 12, no. 4 (2008): 303–24.
———. "The Epistemology of Statistical Evidence." *The International Journal of Evidence & Proof* 15, no. 2 (2011): 117–43. https://doi.org/10.1350/ijep.2011.15.2.373.
Redmayne, Mike. "Exploring the Proof Paradoxes." *Legal Theory* 14, no. 4 (2008): 281–309.
Ross, Lewis D. "Recent Work on the Proof Paradox." *Philosophy Compass* 15, no. 6 (2020). https://doi.org/10.1111/phc3.12667.
Roush, Sherrilyn. "Knowledge, Evidence, and Naked Statistics." In *Externalism about Knowledge*. Oxford University Press, forthcoming.
Smith, Martin. "When Does Evidence Suffice for Conviction?" *Mind* 127, no. 508 (2018): 1193–1218. https://doi.org/10.1093/mind/fzx026.
Southall, Ashley. "A Black Man Spent 7 Years in Prison. Then a Court Changed the Rules on Racial Bias." *The New York Times*, March 18, 2019, sec. New York. https://www.nytimes.com/2019/03/18/nyregion/cross-race-identification-witness.html.
Thomson, Judith Jarvis. "Liability and Individualized Evidence." *Law and Contemporary Problems* 49, no. 3 (1986): 199–219. https://doi.org/10.2307/1191633.
Tribe, Laurence H. "Trial by Mathematics: Precision and Ritual in the Legal Process." *Harvard Law Review* 84, no. 6 (1971): 1329–93. https://doi.org/10.2307/1339610.
Twining, William. "Debating Probabilities." *The Liverpool Law Review* 2, no. 1 (1980): 51–64. https://doi.org/10.1007/BF03186639.
Wasserman, David T. "The Morality of Statistical Proof and the Risk of Mistaken Liability Decision and Interference Litigation." *Cardozo Law Review* 13, no. Issues 2-3 (1991): 935–76.
Wilson, John, Kurt Hugenberg, and Michael Bernstein. "The Cross-Race Effect and Eyewitness Identification: How to Improve Recognition and Reduce Decision Errors in Eyewitness Situations." *Social Issues and Policy Review* 7 (2013). https://doi.org/10.1111/j.1751-2409.2012.01044.x.
Woocher, Fredric D. "Did Your Eyes Deceive You--Expert Psychological Testimony on the Unreliability of Eyewitness Identification Note." *Stanford Law Review* 29, no. 5 (1977): 969–1030.
Wright, Richard. "Causation, Responsibility, Risk, Probability, Naked Statistics, and Proof: Pruning the Bramble Bush by Clarifying the Concepts." *Iowa Law Review* 73 (1988): 1001–77.

The Temporal Disproportionality of the Death Penalty

Many of those who are sentenced to death spend decades awaiting execution. The modern practice of the death penalty thereby includes both long-term incarceration and execution. In light of this long-term incarceration, I present a new argument against capital punishment. For the purposes of this paper, I grant my opponent the basic commitments of a pro-death-penalty retributivist, including that a person can deserve death. Even with these commitments, I argue that capital punishment, unlike other forms of standard punishment, is inherently disproportional given the types of harm involved. First, I contend that there is a necessary tradeoff between the need for a thorough appellate process that ensures desert and the disproportional harm experienced by those on death row. Second, I present some unique proportionality challenges inherent to using death as a penalty. I conclude with a proportionality dilemma for the pro-death-penalty retributivist.

Introduction

On February 3, 2016, Brandon Jones was executed in the United States. Jones was first tried and sentenced in October 1979. He spent 36 years and four months on death row between his conviction and execution.[230]

At the time, Jones was believed to have served the longest time incarcerated between conviction and execution in the US. Jones' long stay on death row, however, is far from an anomaly. As of 2019, the 2,570 people on death row in the US had spent an average of 22 years awaiting execution.[231] Jones and many other condemned people experience both long-term incarceration and execution.

Considering this long-term incarceration, I develop and present a new set of challenges for pro-death-penalty retributivists. Retributivist justifications of the death penalty underlie much of the philosophical pro-death-penalty literature.[232] Though there are different versions of retributivism,

[230] Segura, "Facing Execution at 72, Georgia's Oldest Death Row Inmate Exposes Death Penalty's Racist Roots." All other references to this case are from this source unless indicated otherwise.
[231] Snell, "Capital Punishment, 2019 - Statistical Tables."
[232] See, e.g., Gerstein, "Capital Punishment-"Cruel and Unusual"?"; Sorell, "Aggravated Murder and Capital Punishment"; Flanders, "Time, Death, and Retribution"; Kershnar, "A Defense of Retributivism"; Davis, "How to Make the Punishment Fit the Crime"; Davis, "Harm and Retribution"; Davis, "A Sound Retributive Argument for the Death Penalty"; Hsu, "Does Communicative Retributivism Necessarily Negate Capital Punishment?"; Roark, "Retribution, the Death Penalty,

there are two key principles at the core of all retributivist theories of punishment: desert and proportionality.[233] The retributivist's pro-death-penalty stance is roughly as follows: when a person commits a very serious crime, the person deserves the proportional punishment of death.

Many philosophers have argued against the death penalty by attacking the retributivist's pro-death-penalty stance.[234] In this paper, however, I grant the pro-death-penalty retributivist the key principles of retributivism, as well as the stance that people can deserve death. I also grant the premise that the state can know, to the extent necessary, that a person deserves death. Yet, even if one concedes all of these premises, a set of challenges remains.

In this paper, I argue that capital punishment, unlike incarceration and other forms of standard punishment, is inherently disproportional given the types of harm involved. In §1, I provide an overview of the retributivist's commitments. In §2, I argue that there is a problematic tradeoff between the desert-based need for a thorough appellate process and the disproportional harm experienced by the person incarcerated on death row. I raise and address some objections in §3. In §4, I demonstrate how the appellate process creates disproportionality within a given system of punishment. In §5, I demonstrate some unique challenges inherent to using death as a punishment. I end with a final proportionality dilemma for the pro-death-penalty retributivist.

To demonstrate these challenges, I primarily draw on examples from the US. The practice of capital punishment in the US includes a thorough appellate process. All of the arguments in this paper would similarly apply to any jurisdiction with a death penalty and an appellate process.[235] If a jurisdiction had the death penalty but no appellate process, however, only a subsection of these arguments would apply.[236]

---

and the Limits of Human Judgment"; van den Haag, "The Ultimate Punishment." For a more general overview, see Christopher, "Deterring Retributivism: The Injustice of 'Just' Punishment," 845–47.

[233] For an overview of nine forms of retributivism, see Cottingham, "Varieties of Retribution."

[234] For desert-based arguments: Caruso, "Free Will Skepticism and Criminal Behavior"; Caruso, "Justice Without Retribution"; Alexander, "Retributivism and the Inadvertent Punishment of the Innocent"; Brooks, "Retributivist Arguments against Capital Punishment"; Brooks, "Retribution and Capital Punishment"; Dolinko, "Foreword"; Halvorsen, "Is It Better That Ten Guilty Persons Go Free than That One Innocent Person Be Convicted"; Lempert, "Desert and Deterrence"; Shafer-Landau, "Retributivism and Desert." For proportionality-based arguments: Bedau, "The Minimal Invasion Argument Against the Death Penalty"; Finkelstein, "Death and Retribution"; Roberts-Cady, "Against Retributive Justifications of the Death Penalty"; McDermott, "A Retributivist Argument against Capital Punishment"; Yost, "What's Wrong with Differential Punishment?"; Yost, *Against Capital Punishment*; Lenta and Farland, "Desert, Justice and Capital Punishment."

[235] Other countries with capital punishment and appellate oversight include Barbados, Japan, Singapore, Thailand, Vietnam, Belarus, Bahrain, and Sudan. See "Death Sentences and Executions 2018."

[236] Though most countries with the death penalty technically have appellate oversight, UN experts have found that the appellate processes in some countries are ineffective and thus do not provide real oversight. See, e.g., Rehman et al., "Iran."

§1. Defining Retributivism

Retributivism itself is hard to pin down, since the doctrine has evolved over time.[237] At its core, retributivism can be understood as a justification of punishment based on the desert of the person who committed a crime. If the person deserves punishment for the crime committed, then it can be permissible, or even required, for the state to impose punishment.

Some versions of retributivism understand desert as the only element that justifies punishment. According to those views, punishment is justified because and only because the person deserves to be punished. For other versions of retributivism, desert is a necessary but not sufficient condition for justified punishment. Ultimately, retributivism refers to views on which punishment is justified only if and only to the extent that the person deserves to be punished.

The desert-centered justification for punishment gives rise to a basis for determining the amount of punishment due: people may be punished to the extent that they deserve punishment.[238] Desert can be understood as having three key components: desert subject (S), desert object (O), and desert basis (B).[239] The relation between the three can be understood as follows:

S deserves O by virtue of B.

In cases of punishment, the desert subject is the person who committed the crime. The desert object and the desert basis, however, can be understood in a variety of ways.

In the context of punishment, both the desert basis and the desert object are framed in negative terms.[240] Retributivists understand the desert basis differently. Some argue that the desert basis is the subject's culpability.[241] Others take the desert basis to be the subject's wrongdoing, culpability, or blameworthiness.[242] The desert basis could reflect the resultant harm of the subject's action, or it could only reflect the intentions of the subject.[243] Legal scholars commonly refer to the desert basis as the "gravity" or "seriousness" of the crime.[244] Similarly, retributivists disagree about the

---

[237] See, e.g., Cottingham, "Varieties of Retribution."
[238] See, e.g., Sigler, "Private Prisons, Public Functions, and the Meaning Punishment"; Kleinig, *Punishment and Desert*; Shafer-Landau, "Retributivism and Desert"; Alexander and Ferzan, *Crime and Culpability*.
[239] See, e.g., Sher, *Desert*; Pojman and McLeod, *What Do We Deserve?*
[240] The desert basis and desert object are not necessarily negative. For instance, if the desert object is happiness, then a positive desert basis would mean that an increase in happiness is due, while a negative desert basis would mean that a decrease in happiness is due.
[241] See, e.g., Alexander, "Proportionality's Function"; Alexander and Ferzan, *Crime and Culpability*.
[242] See, e.g., Kadish, "The Criminal Law and the Luck of the Draw"; Kolber, "Against Proportional Punishment."
[243] See, e.g., Gardner, "Crime: In Proportion and in Perspective," 222.
[244] See, e.g., Duff, *Punishment, Communication, and Community*; Boonin, "The Appeal to Necessity"; Husak, "Already Punished Enough"; Hart, "Postscript: Responsibility and Retribution."

desert object. The desert object can be understood as suffering, pain, harm, unhappiness, or the severity of the punishment.[245]

Importantly, both the desert basis and the desert object come in degrees. Regardless of how one interprets the desert basis, there will be instances in which people have a greater desert basis or a lesser desert basis. If the desert basis is culpability, for instance, there are cases in which people are more or less culpable. Alternatively, if the desert basis includes harm, there will be instances with more harm and instances with less harm. The desert basis, therefore, is on a gradient. Similarly, the desert object comes in degrees. If the desert object is well-being, for instances, a person's well-being can be marginally reduced (e.g., fine) or massively reduced (e.g., liberty stripped for many years). The same can be said for the other possibilities; harm, suffering, and the severity of punishment can all be inflicted in degrees.

The relationship between the desert basis and the desert object is a proportional one. Roughly speaking, the proportional relationship can be understood as follows: as the desert basis increases, the desert object also increases. Conversely, as the desert basis decreases, the desert object similarly decreases. Proportionality plays a central role in both determining punishment and limiting punishment. Across the board, retributivists are committed to some version of proportionality. Anthony Duff, for instance, states that proportionality is "intrinsic to any version of retributivism."[246] Thus, I take the term "proportional" to refer to the proper fulfillment of a person's desert object.

Thus, retributivists across the board are committed to both desert and proportionality. Importantly, the two principles are often used to justify the death penalty. Though not all retributivists support the death penalty, many death penalty supporters point to retributivist justifications of the practice. David Dolinko contends that the strongest retributivist argument for the death penalty relies on a general appeal to desert and proportionality:

> We may be unable to compare the gravity of the crime and the severity of the punishment with mathematical precision but we can tell when the two are grossly out of proportion. In particular, we can see that the harm done by killing a person—the permanent and utter annihilation of that individual, the total annulment of all his rights—is so complete and unique

[245] See, e.g., Bagaric and McConvill, "Giving Content to the Principle of Proportionality."
[246] Duff, *Punishment, Communication, and Community*, 132. Mary Sigler similarly contends that "retributivism…is centrally concerned with the imposition of punishment in proportion to an offender's moral desert." Sigler, "Private Prisons, Public Functions, and the Meaning Punishment," 38.

> as to be wholly incommensurable with any imposition, no matter how severe, that yet leaves its victim a living, conscious, sentient subject of experiences and possessor of rights.[247]

Thus, proportionality and desert lie at the heart of retributive justifications of the death penalty.

In theoretical terms, if a person deserves death, then death fulfills the person's desert object. In practical terms, death would be the deserved punishment for murder, for instance. This assumption seems fairly uncontroversial among pro-death-penalty retributivists.[248] Indeed, the application of proportionality is often considered to be at its clearest in the context of the death penalty: if A kills B, then the proportional punishment for A is death.

A common line of objection against pro-death-penalty retributivist is that the state cannot know, to the extent necessary in order to impose the death penalty, that the person deserves execution. Many people commonly point to the fact that innocent people have historically been executed. The pro-death-penalty retributivist would respond that, even though the state has made mistakes in the past, the state can know that a person deserves death to the extent necessary to impose capital punishment. Though the state may not be able to reasonably believe that immediately, the state can arrive at such a belief with the proper safeguards in place, such as a review process of convictions.

Let us grant the pro-death-penalty retributivist two assumptions:

(1) A person can deserve death.

(2) The state can believe that a person deserves death to the extent necessary to impose death as a penalty.

With those assumptions in place, I raise a number of concerns regarding capital punishment for the pro-death-penalty retributivist. I argue that the death penalty, unlike incarceration, has features that make it extremely difficult, if not impossible, to impose as a proportional penalty. Notably, my view is the precise opposite of the commonly-held stance that the death penalty for killing another person is the clearest application of proportionality.

§2. Pre-Execution Harm

In this section, I establish the first problem for the pro-death-penalty retributivist. The problem is as follows: a thorough appellate process serves as an important check to ensure that the person indeed deserves the death penalty. The length of the appellate process, however, does not correspond to the

[247] Dolinko, "Foreword," 599–600.

[248] See, e.g., Gerstein, "Capital Punishment-"Cruel and Unusual"?"; Berns, "For Capital Punishment"; van den Haag, "The Ultimate Punishment."

desert of the person on death row. The long period of incarceration during which the appellate process runs its course introduces disproportionality. Thus, the pro-death-penalty retributivist faces a tradeoff between ensuring desert and imposing a proportional punishment.

This section proceeds in two parts. First, I explain why the pro-death-penalty retributivist would support a thorough appellate process to ensure desert. Second, I demonstrate that long-term incarceration while awaiting execution is extremely harmful. Since the harm is determined by the length of the appellate process, and not by the desert of the person on death row, the harm experienced on death row leads to disproportional punishment.

2.1 Appellate Process

First, it is important to ensure that the person receiving the death penalty indeed deserves the death penalty. As noted above, desert is considered the defining feature of retributivism. For a retributivist, desert is a necessary condition for permissible punishment. If a person does not deserve the punishment, then the punishment is never justified. Given that a punishment is justified only if and only to the extent that the person deserves to be punished, a retributivist would certainly support ensuring—to the fullest extent possible—that the person is indeed deserving of the given punishment.

There are a variety of ways that the appellate process is used to ensure that people are indeed deserving of punishment. First and foremost, a person may attempt to demonstrate her innocence. Indeed, innocent people frequently end up on death row. Since the US restarted capital punishment in 1976, 185 people have been exonerated while on death row through the appellate process.[249] Indeed, for every 8.3 people executed by the US, one person has been exonerated through the appellate process.[250] Thus, a thorough and lengthy appellate process dramatically reduces the number of innocent people executed by providing people sentenced to death a chance to appeal the decision.

Second, people who are sentenced to death may also use the appellate process to demonstrate that they do not deserve capital punishment. For instance, in Atkins v. Virginia (2002), the US Supreme Court found that it was unconstitutional to execute minors or those who are intellectually disabled.[251] Even if a minor or an intellectually disabled person did in fact commit the relevant crime, the court found that such a person is not fully responsible for her actions. If the person is not fully responsible for her actions, the court reasoned, then she does not deserve the death penalty. Importantly, the

[249] Death Penalty Information Center, "DPIC Adds Eleven Cases to Innocence List, Bringing National Death-Row Exoneration Total to 185."
[250] Death Penalty Information Center.
[251] Atkins v. Virginia, 536 U.S. 304 (2002).

person can still face a serious, long-term sentence; she just cannot be sufficiently deserving of capital punishment. The appellate process is frequently used to determine whether the person sentenced to death has an intellectual disability. Indeed, there are a variety of cases in which death sentences have been commuted to life-without-parole sentences on the basis of the person's intellectual disability.[252]

Third, people who are sentenced to death may also use the appellate process to demonstrate that the expected suffering is too harsh relative to what is deserved, even though the person is fully responsible for the crime. Russell Bucklew, for instance, was convicted and sentenced to death.[253] During the appellate process, he raised the issue that a lethal injection would constitute "cruel and unusual" punishment because of his congenital medical condition. Due to the unusual condition, doctors expected that the lethal injection would cause him to hemorrhage and die by choking on his own blood. The argument made during the appellate process contended that, although he may deserve death, Bucklew did not deserve the excruciating death that would result from the combination of the lethal injection and his medical condition. (His request was to be executed by a different method.) The appellate process therefore also provides an avenue for death row detainees to raise arguments that the level of pain would be undeserved.

Given that there are a variety of desert-related claims made during the appellate process, the process helps ensure that people who are undeserving of death (e.g., due to either their innocence or their lack of full responsibility) are not executed. Determining desert is crucial for all types of punishment. For the death penalty, however, it is of extra importance to ensure that the person deserves the punishment, since it is impossible to retract. If a person is undeservedly incarcerated for many years, the state cannot give back those years lost but the state can compensate the person. If the state discovers that an undeserving person was executed, however, the state cannot take remedial steps, since the person is dead. Thus, it seems that the pro-death-penalty retributivist would certainly support a thorough appeals process in order to ensure that those who are punished are indeed guilty and deserving.

A thorough appellate process, however, takes time. In the US, after a person is sentenced to death following a criminal trial, the defendant may begin the nine-step appellate process.[254] This process can take decades, especially since certain types of arguments can only be raised later in the process.[255] The length of the process may be impacted by a variety of factors, including the speed of

[252] Johnson, Blume, and Van Winkle, "Atkins v. Virginia at Twenty."
[253] Totenberg and Montanaro, "Supreme Court Closely Divides On 'Cruel and Unusual' Death Penalty Case."
[254] Stevenson, "The Politics of Fear and Death," 717.
[255] Coleman, "Litigating at the Speed of Light."

the courts, the responsiveness of judges, the complexity of the legal arguments, and the priorities of local prosecutors.[256] Only once a defendant has exhausted this nine-step appellate process is the litigation considered final.[257] After a defendant exhausts the appellate process, the state may issue a death warrant. Once a death warrant is issued, more litigation may occur, since there are certain types of concerns that can only be raised after a warrant is in place. But death warrants are not issued at a regular pace. Rather, they seem to respond to external factors, including political whims.[258] As a result, some defendants exhaust the appellate process and simply wait on death row for years. Since the reinstatement of the death penalty in 1976, the average time spent on death row pre-execution has steadily increased. In 1985, the average time spent on death row was five years and 11 months.[259] As of 2019, the average time spent on death row is 22 years.[260]

Thus, a pro-death-penalty retributivist would support a thorough appellate process, even though it takes time. The challenge with the appellate process, however, is that throughout the appellate process, defendants are considered "condemned" as they await their executions on death row. As a result, they suffer from a serious (and largely overlooked) harm, which I address in the coming section.

2.2 Long-Term Incarceration

I argue that long-term incarceration while anticipating execution is itself an extreme harm that warrants the independent attention of the pro-death-penalty retributivist.

Long-term incarceration is generally an extremely harsh and unpleasant experience, one that is inflicted in response to a variety of other serious crimes. Both those who are sentenced to death and those who are serving long sentences experience physical and emotional harm. Furthermore, many psychology studies have found that long-term incarceration takes a severe toll on an individual's mental health.[261] Moreover, in the US, death row detainees are frequently held in solitary confinement

---

256 Stevenson, "The Politics of Fear and Death."
257 Stevenson, 717.
258 The most high-profile example of political influence on executions was former President Bill Clinton's return to Arkansas in the middle of his campaign to oversee an execution as the governor to send a message that he was tough on crime. See Fournier, "The Time Bill Clinton and I Killed a Man." A recent example of political influence on executions was former President Trump's restart of federal executions after a 20-year hiatus around the time that he started his re-election campaign. As soon as President Biden was elected, the moratorium on executions was put back in place. See Wise, "The Justice Department Is Pausing Federal Executions After They Resumed Under Trump." For an explanation of how "political will" influences state executions, see, e.g., Steiker and Steiker, "A Tale of Two Nations."
259 ProCon, "Death Row Inmates."
260 See Snell, "Capital Punishment, 2019 - Statistical Tables."
261 See, e.g., Bonta and Gendreau, "Reexamining the Cruel and Unusual Punishment of Prison Life"; Fogel, *Developing through Relationships*; McCorkle, "Personal Precautions to Violence."

for years. Psychologist Hans Toch (2007) conducted a large-scale study on the effects of imprisonment and isolation on those incarcerated in New York State correctional facilities.[262] He found that there is a critical "distinction between imprisonment, which is tolerable, and isolation, which is not."[263]

The primary harm experienced by death row detainees, however, often comes from the psychological anticipation of death, not from the conditions themselves. Psychologist Amy Smith argues that the psychological harm of awaiting execution is unique.[264] Smith acknowledges that those who are incarcerated without a death sentence and those who are incarcerated while awaiting executions may be facing similar physical conditions. Yet, those who await execution experience an additional psychological harm, the constant anticipation of one's own death. Smith points out, moreover, that there is an important distinction between the experience of awaiting execution on death row and the experience of awaiting natural death from terminal disease. She writes that "the permanent fear of being forced to die at the hands of another is an exquisite psychological torture that creates severe emotional, mental, and also physical suffering."[265]

The California Supreme Court has also recognized that incarcerating people while they anticipate execution is a form of torture:

> The cruelty of capital punishment lies not only in the execution itself and the pain incident thereto, but also in the dehumanizing effects of the lengthy imprisonment prior to execution during which the judicial and administrative procedures essential to due process of law are carried out. Penologists and medical experts agree that the process of carrying out a verdict of death is often so degrading and brutalizing to the human spirit as to constitute psychological torture.[266]

Similarly, Albert Camus writes: "As a general rule, a man is undone by waiting for capital punishment well before he dies. Two deaths are inflicted on him, the first being worse than the second."[267] Across decades and jurisdictions, there is a recognition that the psychological anticipation of execution may be worse than death itself.

Let us return to the premise that we granted the retributivist at the beginning of the paper, namely that death can be the proportional punishment to A's desert basis. By the time A's execution date arrives, however, A has already suffered immensely. As a result, her desert object is different than

---

[262] Toch, *Men in Crisis.*
[263] Toch, 54.
[264] Smith, "Not 'Waiving' But Drowning: The Anatomy of Death Row Syndrome and Volunteering for Execution."
[265] Smith, 250.
[266] People v Anderson at 894–95.
[267] Camus, "Reflections on the Guillotine," 156.

it was at the time of sentencing. Even if she originally deserved death, the long-term incarceration constitutes its own suffering, such that inflicting death on the execution day may be disproportional relative to A's desert object at that point. Moreover, the degree of suffering largely depends on factors that are external to A's desert. As a result, a severe harm, which is completely non-responsive to the desert basis, is inflicted before execution. If a person's desert object diminishes as she is harmed, then the proportional treatment would change over time. If so, this poses a challenge to the permissibility of execution many years after conviction.

To demonstrate this challenge, let us return to the case of Brandon Jones from the beginning of the paper. Jones was born in 1943 and was convicted of a crime he committed as a 36-year-old in 1979. After 36 years and four months on death row, Jones was executed in 2016 at the age of 72. Jones spent *half* of his life on death row.

Let us suppose that Jones deserved death when he was sentenced to death in 1979. If so, then his desert object, x, could be fulfilled through the death penalty. As Jones suffered on death row, however, his desert object would decrease. Thus, suppose that after 36 years on death row, Jones' desert object is now a fraction of x. If so, the death penalty would be too harsh of a penalty relative to his desert object in 2016. If a person spends a matter of days on death row, the diminishment of the desert object through incarceration may not be sufficient to make the death penalty a disproportionately harsh punishment. Yet, a process that spans years is vulnerable to this type of disproportionality.

Importantly, this challenge is unique to the death penalty, since what is thought to be the primary penalty is backloaded in the sentence. The diminishment of desert objects over time explains why standard sentences end. If a person is sentenced to 30 years, the diminishment of the person's desert object is reflected in the remainder of the sentence continuously decreasing. Suppose that a person is sentenced to 30 years to fulfill her desert object which is 30x. After 15 years, the person's desert object may have diminished to 15x because of the years spent incarcerated. The person at that point only faces 15 more years, not the original 30. Then, after 30 years, the person's desert object may be at zero and thus no more incarceration is deserved.

The lengthy appellate process poses a proportionality problem if desert objects diminish over time as a person suffers. Even if a person deserved death when sentenced, she may not deserve death by the time an execution date arrives. Thus, the long-term incarceration before execution may make the death penalty disproportional when the execution date arrives.

§3. Objections

So far, we have established that a thorough death penalty process is necessary but lengthy. While the process happens, the person suffers while awaiting execution. The harm experienced on death row changes the degree of harm required to properly fulfill the desert object. Thus, when the person faces execution, the person may not be receiving the proper proportional penalty. In this section, I raise four objections on behalf of the pro-death-penalty retributivist: (1) Deserve Death+, (2) Shorten or Improve, (3) Not Punishment, and (4) Not Harm. I will address each objection in turn.

3.1 Deserve Death+ Objection

First, a pro-death-penalty retributivist may object as follows:

> **Deserve Death+**: People who are sentenced to death deserve more than just the punishment of death. Rather, long-term incarceration and death better fulfills their desert objects.

A pro-death-penalty retributivist could argue that, in many cases, the person deserves a fate worse than death. In some cases, a person may have killed multiple people. Alternatively, a person may have harmed and killed the victim. A pro-death-penalty retributivist may argue that only execution is not a sufficient punishment, and thus long-term incarceration and execution gives the person closer to what is deserved.

I respond that one has to assume that a person deserves more than death to even begin assessing the disproportionality introduced by a thorough appellate process. If a person deserves only death and absolutely no more, then even just one day would introduce disproportionality. Rather, this discussion already must assume that the person could deserve more than death.[268] Yet, even when a person deserves more than death, she may still be receiving a disproportional punishment, since the length of suffering is not responsive to her desert.

3.2 Shorten or Improve Objection

Second, a pro-death-penalty retributivist may object as follows:

> **Shorten or Improve**: The tradeoff between ensuring desert and proportionally harming people on death row only exists because the appellate process is so long and conditions on

[268] On a practical note, since the death penalty necessarily includes long-term incarceration, the pro-death-penalty retributivist would likely only support the death penalty for especially heinous or severe crimes. This, however, is already a major concession, since many people who are sentenced to death have not actually committed the most heinous crimes. This would entail that the standard practice of the death penalty is often disproportional.

death row are so miserable. Shortening the appellate process or improving the conditions on death row would resolve the issue.

There are two suggested solutions in this objection: shorten the appellate process and improve the conditions. I will address each in turn.

First, one may point out that the appellate process does not need to be as long as it is in the US. Indeed, it seems unnecessary for the process to last over two decades on average. If the appellate process was dramatically shortened, then perhaps its lack of responsiveness to the desert of the death row detainee would be less important.

I have two responses. First, while the appellate process could (and should) certainly be reformed to be less affected by external factors such as political whims, a thorough process is still needed to ensure the presence of desert. Indeed, the thorough and lengthy appellate process in the US has yielded 185 death row exonerations.[269] While it may be possible to shorten the process to some extent while maintaining its desert-checking capabilities, excessively truncating the process could lead to the execution of undeserving people—an outcome that a pro-death-penalty retributivist would certainly oppose. Thus, there is a necessary tradeoff between the length and accuracy of the process.

Second, even if one could reduce the length of the appellate process and maintain the same level of accuracy, the appellate process would still take some amount of time. Even if the appellate process was shortened to only a few years, that amount of time is still substantial, especially when spent on death row. Indeed, the British court that served as the highest appellate court for countries in the Caribbean Commonwealth found that execution after five years on death row was unconstitutional because it was "inhumane and degrading."[270] If a person reached five years on death row, the court would automatically commute the person's death sentence to life in prison. Similarly, in 1890, the US Supreme Court found that long delays before execution could result in "one of the most horrible feelings" and "immense mental anxiety amounting to a great increase in the offender's punishment."[271] The long delay in that case was four weeks. Moreover, the length of the appellate process would still be dependent on factors external to the person's desert, and thus the challenge would remain.

---

[269] Death Penalty Information Center, "DPIC Adds Eleven Cases to Innocence List, Bringing National Death-Row Exoneration Total to 185."
[270] Bowcott and Wolfe-Robinson, "British Court to Rule on Death Sentences for Two Trinidad Murderers."
[271] Medley, Petitioner, 134 U.S. 160 (1890) at 172.

The second suggestion in the objection is that the physical conditions of pre-execution incarceration could be considerably improved. The use of solitary confinement could be abolished, for instance, and people could be incarcerated with the general prison population instead.

While I certainly support measures taken to reduce the suffering experienced between sentencing and execution, the primary harm experienced on death row is the anticipation of execution. As psychologists, philosophers, and even the US Supreme Court have acknowledged, the suffering on death row is primarily psychological, which cannot be ameliorated by better physical living conditions. Anticipation of execution necessarily follows receiving a death sentence. (The only exception is when people are too mentally ill to understand that they face execution. That of course has its own major issues.)

Thus, shortening the appellate process or improving conditions cannot fix the problem. Even if the appellate process was dramatically shortened, the time on death row still would be significant and non-responsive to desert. Moreover, between conviction and execution, the person necessarily suffers from the anticipation of execution, even if physical conditions are improved. Thus, during a thorough appellate process, the person on death row necessarily suffers from the anticipation of execution, forcing a tradeoff between a needed appellate process and disproportional harm.

3.3 Not Punishment Objection

Third, a pro-death-penalty retributivist may object as follows:

> **Not Punishment**: The long-term incarceration while a person is awaiting execution should not be considered part of the deserved treatment, since it is not punishment.

A proponent of this objection would argue that the long-term incarceration is an opportunity to appeal the sentence and thus should not be part of the calculation of the desert object.

In order to respond to the objection, it is important to untangle the assumptions it makes. The objection assumes (a) that long-term incarceration on death row is not punishment, and (b) that only something that is "punishment" should be considered when imposing what is deserved. I will respond to each assumption in turn.

In response to (a), I argue that the long-term incarceration on death row is punishment. The same argument that I previously presented regarding pre-trial detention applies here as well, except that incarceration on death row even more obviously constitutes punishment. I will quickly run through the argument to explain how it applies here.

First, incarceration on death row meets the five criteria of H. L. A. Hart's definition of punishment.[272] Long-term incarceration certainly involves "pain or other consequences normally considered unpleasant" (i). A person on death row is an "actual or supposed offender" who has been convicted of a serious crime, which is "an offence against legal rules" (ii, iii). Death row incarceration is enforced by prison guards, who are "human beings other than the offender" and are an "authority constituted by a legal system against which the offence is committed" (iv, v). In the same way that a person is punished *for a* crime, a person is on death row *for* murder or another crime.

Second, incarceration on death row can be punishment even if it serves other purposes. As we have seen, pre-trial punishment and post-trial incarceration may serve to incapacitate a person, for instance. Similarly, death row may serve to ensure that the person is available for execution. Yet, the presence of these other aims does not negate the possibility that the treatment is punishment. Even if the state claims that incarceration on death row is not punishment, it can still constitute punishment. As argued in Chapter 2, there is good reason to ignore the state's claims regarding whether something is intended as punishment. For instance, the state may want to remove the constitutional protection against "cruel and unusual" punishment, since death row likely qualifies as such and thus would be unconstitutional if recognized as punishment. Moreover, as demonstrated in Chapter 2, actions speak louder than words. The conditions on death row are extremely harsh, and the state compensates death row exonerees, just as it compensates those exonerated from standard prison sentences.[273]

Third, death row incarceration also meets the two requirements of Yaffe's Norm Shift Theory definition of punishment-by-nature.[274] First, death row incarceration is a required meaningful reduction of the person's normative freedom since, once sentenced to death, it is illegal for that specific person to be elsewhere. Second, death row incarceration is an exercise of the state's punitive power, which Yaffe takes to mean that there was a finding that the person committed a crime. Unlike people in pre-trial detention, people on death row have in fact been convicted. That conviction straightforwardly fulfills Yaffe's definition of a finding.

Finally, incarceration on death row is even more obviously punishment than pre-trial punishment. It is post-conviction and only imposed on people who have been found guilty, just like standard punishment. People are incarcerated in the same facilities, namely prisons, as people who are serving standard punishment sentences. Moreover, unlike the selective application of pre-trial

[272] Hart, *Punishment and Responsibility*, 4–5.
[273] The US federal government compensates exonerees $50,000 per year incarcerated and an extra $50,000 per year spent on death row. See Slifer, "How the Wrongfully Convicted Are Compensated for Years Lost."
[274] Yaffe, "The Norm Shift Theory of Punishment."

punishment, death row incarceration is systematically imposed on everyone who is sentenced to death. Thus, incarceration on death row is punishment.

Assumption (b) of the objection states that only punishment should be considered when imposing what is deserved. Even if we granted a proponent of the objection that incarceration on death row is not punishment, I argue that the pro-death-penalty retributivist would still support taking the long-term incarceration into consideration.

First, in practice, the severity of the punishment inflicted is often mitigated if the person has experienced other harm that is not officially called punishment. One of the clearest examples of this is the practice of crediting time served in pre-trial detention. Though I elsewhere argue that pre-trial detention is and should be considered punishment, the US Supreme Court maintains that it is not punishment.[275] Yet, post-trial sentences are still reduced when a person has been incarcerated in pre-trial detention.

Second, the severity of punishment is also mitigated mid-sentence if there is a concern that the person will be harmed more than deserved as a result of circumstances that are not punishment. For instance, during the start of the COVID-19 pandemic, many people were released before the end of their sentences from US prisons due to medical conditions that meant that catching COVID-19 could lead to major physical harm or even death. To see how this is analogous, suppose that a person had five years remaining in their sentence. When the pandemic started, the state had a choice between imposing the remaining five years, which inevitably comes with a risk of death, or under-punishing the person by sending her home. Similarly, when a person is sentenced to death, the state can choose between imposing the death sentence, which inevitably comes with long-term incarceration, or under-punishing the person. The death sentence is analogous to the remaining five years of the sentence, and the risk of death by infectious disease is analogous to the long-term incarceration. In the context of the pandemic, the state often chose to under-punish as a result of the risk posed by the infectious disease, even though infection is certainly not punishment. Thus, the state mitigates punishments as a result of things that are not punishment. Similarly, it would be consistent for the state to choose to under-punish because of the long-term incarceration, even if the long-term incarceration is not punishment.

Third, retributivism, as construed in the literature, supports a holistic approach to determining the desert object, which entails accounting for non-punishment. Ultimately, retributivists care about

[275] United States v. Salerno 481 U.S. 739 (1987). For my argument that pretrial detention is punishment, see Chapter 2 of this thesis.

the desert object, which could be suffering, pain, harm, or unhappiness.[276] Let us say that the desert object is suffering. If so, then punishment is just the term used for the infliction of suffering to fulfill a desert object. Retributivists, therefore, do not care whether the desert object is fulfilled through suffering that is called "punishment" or through suffering that is called something else. As Adam Kolber explains, retributivists believe that "offenders deserve to suffer, even when that suffering is not specifically intended as punishment for an offense."[277] Even if the long-term incarceration is not punishment, it would still be considered part of the desert object. Thus, the objection fails.

3.4 No Harm Objection

Finally, a pro-death-penalty retributivist may object as follows:

> **No Harm**: The long-term incarceration while awaiting execution is not a harm, since the alternative would have been execution. The extra days of life make those sentenced to death better off than they would have been otherwise.

Without the appellate process, the person would have been executed. In almost all circumstances, additional days of life—even if life is challenging—is considered a desirable and good thing.

I argue that the additional days of life on death row do not necessarily make a person better off than they would be otherwise. A person is only better off with an extended life if the extra days are days of a life worth living. Though it is difficult to conclusively determine whether the lives of people on death row are worth living, I contend that there is reason to believe that a life awaiting execution may not be a life worth living. As Camus put it, life awaiting execution is a "death" which is "worse than death itself."[278]

A pro-death-penalty retributivist may reply that the fact that people choose to engage in the appellate process demonstrates that they think their lives are worth living. If they did not think their lives were worth living, then they would have opted for immediate execution.

I have two responses to this argument. First, many of those who choose to proceed with the appeals process might not believe that they are simply choosing long-term death with inevitable execution. Instead, those on death row may believe that they will not ultimately be executed, or at least that there is some substantial chance that their executions will not happen.[279] If so, death row detainees

---

[276] See, e.g., Bagaric and McConvill, "Giving Content to the Principle of Proportionality."
[277] Kolber, "Against Proportional Punishment," 1155.
[278] Camus, "Reflections on the Guillotine," 156.
[279] Death row detainees may not ultimately be executed due to exonerations, sentence commutations, execution moratoriums, or death penalty abolition.

may actually be choosing long-term incarceration with a chance of avoiding execution over guaranteed execution. Furthermore, many individuals may proceed with the appellate process, since waiving the right to appeal does not actually guarantee death. There are a variety of external factors, including political will and access to lethal drugs, that can delay executions.[280] Thus, even if a person definitively prefers immediate execution, she may still opt to appeal because waiving one's right to appeal does not guarantee immediate death.

Second, many people do attempt to choose immediate death over life on death row. Focusing on men in the US, David Lester and Christine Tartaro found that approximately 130 per 100,000 people per year die by suicide on death row.[281] Notably, this number excludes those who wanted to end their lives but were unable to do so. The death row suicide rate is more than five times both that of the general population and of the non-death row prison population.[282] Moreover, a sizeable percentage of those executed in the US do waive their right to appeal, opting for execution over life on death row. According to John Blume, the US executed 885 people between 1977–2003.[283] Of those executed, 106 were "volunteers," death row detainees who waived their right to appeal and opted for execution instead.

Even if life on death row is a life worth living, people are still experiencing an immense amount of extra suffering imposed by the state. Though the alternative to extra days of life on death row may be execution, a person can still suffer to different extents during those extra days of life. For instance, consider the following scenario: A kidnaps B and gives B two options—either to be killed now or to be killed later. B opts to be killed later in hopes of not getting killed at all. A then proceeds to torture B. It would be unreasonable for A to then claim that the torture was not harmful since an alternative would have been death for B. Thus, just because someone chooses extra days of life and feels as though those days of life are worth living, does not mean that the person's extra suffering during that period is irrelevant.

In conclusion, incarceration during the appellate process creates a problem for the pro-death-penalty retributivist. On the one hand, it is important to maintain the appellate process to ensure that undeserving people are not executed. On the other hand, the experience of long-term incarceration while awaiting execution may be depriving a person of days of a life worth living. Even if the days of

---

[280] See, e.g., Tabachnick, "Ohio Governor Postpones 3 Executions amid Struggle to Find Lethal Drug Suppliers." For an overview of reasons why executions are delayed, see Kovarsky, "The American Execution Queue."
[281] Tartaro and Lester, "Suicide on Death Row."
[282] According to Tartaro and Lester, the suicide rate is approximately 25 per 100,000 people in the general population and approximately 17 per 100,000 people in the non-death row prison population.
[283] Blume, "Killing the Willing."

life on death row are worth living, the person is still suffering immensely. Thus, the time on death row is itself an extreme harm that creates disproportional punishment.

If, however, someone insisted that the time on death row is not a harm but rather a positive benefit to those on death row, there would still be a proportionality problem—the time on death row is not only giving a person extra days of a life worth living, but it is also reducing the harm of ultimate execution, thus making the overall punishment better on two fronts.

Even so, systematic under-punishment is still a concern for the retributivist. It is important to note that this type of disproportionality—namely treating people more leniently than how they should be treated—seems to be less of a concern for retributivists. Yet, it is still problematic. Kolber, for instance, challenges retributivists by noting that the practice of crediting time served leads to the systematic under-punishment of people detained pre-trial.[284] Thus, the challenge remains for the retributivist—even if death row is a non-harmful, positive experience, the time on death row amplifies the leniency of the punishment by also reducing the harm of the eventual execution.

Furthermore, if people who are systematically under-punished are treated differently, it is still a problem for the retributivist. If the degree of under-punishment is non-responsive to a person's desert basis, then two people with the same desert bases could experience different degrees of harm. Thus, if one of the two is under-punished, there is a proportionality concern for the retributivist. I will demonstrate this substantial proportionality problem for the pro-death-penalty retributivist in the coming section.

§4. Pre-Execution Ordinal Disproportionality

Thus far, I have argued that death row poses serious proportionality challenges for the pro-death-penalty retributivist. The length spent incarcerated pre-execution is dependent on many external factors, including the political will, court backlogs, prosecutorial discretion, and even access to lethal drugs.[285] The fact that the length of incarceration is dependent on factors that are non-responsive to desert leads to disproportionality.

One may object, however, that our working understanding of proportionality is impossible to implement, since there is no way to know precisely how much suffering a person deserves. Rather,

[284] Kolber, "Against Proportional Punishment." In Chapter 2 of this thesis, I contest Kolber's argument that credit for time served poses a proportionality problem.

[285] See, e.g., Kovarsky, "The American Execution Queue"; Stevenson, "The Politics of Fear and Death"; Steiker and Steiker, "A Tale of Two Nations."

perhaps the notion of proportionality is meant to capture a sense of fairness, ensuring that people who are equally blameworthy are treated the same.

Andrew von Hirsch raises this general concern and instead proposes a differentiation between two measures of proportionality: ordinal and cardinal.[286] Von Hirsch describes ordinal proportionality as the notion that punishments must be properly related to one another. Though we may not be able to calculate how much suffering a given person deserves, we can tell if two people who are roughly equally deserving receive roughly the same punishment, which von Hirsch terms "parity" proportionality.[287] Similarly, we can rank the seriousness of crimes (e.g., murder is worse than robbery which is worse than petty theft) and the seriousness of potential penalties (e.g., a 20-year sentence is worse than a 10-year sentence which is worse than a 1-year sentence). With these two sets ordered, we can then map them onto one another so that major crimes receive worse punishments than do minor crimes. Von Hirsch refers to this type of ordinal proportionality as "rank" proportionality. In contrast to both of these forms of ordinal proportionality, von Hirsch describes cardinal proportionality as "nonrelative," meaning that the only components considered are an individual person and her specific crime.[288]

I argue that the death penalty has rampant ordinal proportionality problems, since one component of the death penalty, incarceration on death row, is non-responsive to a person's desert. Punishments will therefore necessarily violate both parity and rank requirements.

Since the length of pre-execution incarceration is non-responsive to a person's desert, punishments will inevitably violate both parity and rank requirements. In order to demonstrate the parity challenge that incarceration during the appellate process poses, suppose that A and B both commit double homicides in 1990 and deserve the exact same punishment. A has a complicated appellate process, and she ends up spending 20 years in prison before her execution in 2010. B, however, has a relatively uncomplicated appellate process, and she is executed in 1994, after spending four years incarcerated pre-execution.

Even though A and B deserve the exact same punishment, A is punished with 20 years of incarceration and death, while B is punished with four years of incarceration and death. The disparity between the two punishments is not related to the crimes committed, nor is it related to the deservingness of the individuals. As von Hirsch explains, the parity component of proportionality

[286] Von Hirsch, "Proportionality in the Philosophy of Punishment."
[287] Von Hirsch 79.
[288] Von Hirsch, 77.

requires that "when offenders have been convicted of crimes of similar seriousness they deserve penalties of comparable severity."[289] Yet, A and B's penalties are not of comparable severity; A's incarceration is five times the length of B's incarceration.[290] Thus, if the length of incarceration is dependent on these external factors, then violations of the parity requirement are bound to occur.

This pattern of parity disproportionality exists in Jones' case. Jones had a co-defendant, Van Roosevelt Solomon. Both Solomon and Jones were sentenced to death in 1979. While Jones was executed in 2016, after 36 years on death row, Solomon was executed in 1985, after six years on death row. Even though they were convicted of the same crime at the same time, they received dramatically different lengths of pre-execution incarceration.

The rank ordering challenge for the pro-death-penalty retributivist is similar. Suppose that C commits a double homicide in 1990. That same year, D kills 50 people in a mass shooting. As von Hirsch explains, rank ordering requires that "[the] relative severity [of a punishment] reflects the seriousness-ranking of the crimes involved."[291] This means that if D's crime is worse than C's crime, then D should be punished more harshly than C.

Yet, suppose that C's appellate process takes 20 years due to complications and that she is executed in 2010. Suppose also that D's appellate process is somewhat simpler, but still lasts 15 years, so D is executed in 2005. C's total punishment is 20 years of incarceration before death, while D's total punishment is 15 years of incarceration before death. If we assume that a longer pre-execution incarceration is worse, then C receives a harsher punishment than D, despite the fact that D's crime is worse.[292] Thus, this case poses a challenge to the rank ordering requirement of ordinal proportionality.

The parity and the rank ordering requirements both create ordinal proportionality challenges for the death penalty. In order to ensure an ordinally proportional scheme, punishments should be primarily responsive to a person's level of desert. If the degree of harm imposed pre-execution is primarily impacted by factors that are non-responsive to desert, then the system will inevitably be disproportional. Thus, even with this understanding of proportionality, the death penalty faces serious proportionality challenges.

---

[289] Von Hirsch, 79.

[290] For the purposes of this argument, I do not take a stance as to which is worse. I only show that there is a disparity between A and B's treatment.

[291] Von Hirsch, "Proportionality in the Philosophy of Punishment," 79.

[292] If one thinks that C's crime is worse than D's crime, then the argument works if one flips C's punishment and D's punishment. If one thinks that D's punishment is harsher than C's punishment, then the argument still works if one flips C's crime and D's crime.

Given the ordinal disproportionality problems introduced by the appellate process, the pro-death-penalty retributivist faces her first dilemma. On the one hand, the pro-death-penalty retributivist could opt to maintain the current practice of capital punishment, which includes a thorough appellate process. The choice, however, entails proportionality violations: people receive either more or less than they deserve, the death penalty may not even be deserved by the time the execution date arrives, and ordinal disproportionality is inevitable. Thus, a pro-death-penalty retributivist should reject this option.

On the other hand, the pro-death-penalty retributivist could opt to eliminate the appellate process. Though this solution would eliminate the disproportionality introduced by the appellate process, it would also mean that many people may receive undeserved punishments. The appellate process allows for important review of original convictions and sentences, thus providing a chance for people to prove their innocence or demonstrate that the death penalty is undeserved on some other ground. Appellate review is especially crucial given the biases, disparities, and mistakes that can lead to wrongful convictions. The choice to eliminate the appellate process, then, raises serious desert concerns. Thus, this option seems similarly undesirable for the pro-death-penalty retributivist. Since both options are problematic, it is hard for the pro-death-penalty retributivist to resolve the disproportionality introduced by a procedure that is needed to ensure desert.

There is one potential route out of this dilemma—there may be a way to standardize the length of incarceration while awaiting execution, such that people would be treated equally and the appellate process would have sufficient time to run its course. As discussed above, anyone who deserves *only* death would not be able to be sentenced to death, since a death sentence necessarily entails a penalty that is more than death. Rather, the death penalty would only be applied to those who deserve a penalty worse than death. Suppose that the appellate process could run its proper course in 10 years. Then, the following sentencing scheme could be proposed: the penalty for a double homicide is 10 years of incarceration and execution, the penalty for a triple homicide is 15 years of incarceration and execution, and so on.

This proposal has a number of both practical and theoretical problems. What happens if the appellate process takes longer than 10 years? If new evidence is uncovered after 9.5 years but needs review, would the appellate process be extended? Theoretically, could a person opt out of the pre-execution years of incarceration? What types of crimes would correspond to these sentences given the immense amount of suffering involved in awaiting one's own execution?

Even if all of these issues could be resolved, this route out of the dilemma takes the pro-death-penalty retributivist straight into a new dilemma. The new dilemma arises from disproportionality problems at the back end, when it comes time for execution. In the coming section, I explain the proportionality problems with the actual execution as opposed to the time spent awaiting execution.

§5. Execution Disproportionality

For now, let us put aside our concerns about pre-execution harm and turn to the harm of execution itself. I argue that executions, unlike incarceration and other standard punishments, pose inescapable standard and ordinal proportionality problems due to the nature of the harm of the punishment.

Though the actual execution may or may not be painful, the primary harm of death is depriving a person of the future life she would have had.[293] Since the harm is primarily the loss of days of a life worth living, the harm of execution will depend on a variety of factors, including the person's age and health. To demonstrate this, let us compare the execution of a 95-year-old to that of a 25-year-old. The 95-year-old likely would have had only a few more years of life to live. The 25-year-old, however, likely would have had many decades of life left to live.

This poses a standard proportionality challenge for a pro-death-penalty retributivist, since people may be harmed more or less than they deserve. First, people may suffer far more or far less than deserved. For instance, if the 95-year-old deserved a lot of suffering, then depriving the 95-year-old of only a few months of life may seem like an under-punishment relative to what is deserved. Alternatively, if the 25-year-old deserved a lesser degree of suffering, then perhaps it would be best to deprive the 25-year-old of 10 years of life, not many decades of life. Since the primary harm involved in execution can vary greatly depending on a person's age, health, and other factors, it may not be possible to impose the deserved degree of suffering through execution.[294]

The harm of execution is impacted by yet another factor that does not correspond to desert: time spent on death row. As a person spends more time on death row, the person necessarily gets older. As each day passes while on death row, the person is deprived of fewer and fewer future days of life when executed. Thus, as the harm of death row increases, the harm of execution decreases. Though time on death row decreases the harm of execution, it is still cumulatively more harmful; the

[293] For some of the central literature on the harms of killing, see, e.g., McMahan, "Death and the Value of Life"; McMahan, "Causing People to Exist and Saving People's Lives"; Millum, "Age and Death"; Lippert-Rasmussen, "Why Killing Some People Is More Seriously Wrong than Killing Others"; Nagel, "Death."

[294] Indeed, the US has executed and attempted to execute people who are terminally ill. See Chammah and Blakinger, "They Are Terminally Ill. States Want to Execute Them Anyway."

harm of extra time on death row far exceeds the small decrease in harm when a person is deprived of fewer days of a future life. The primary harm of death row is being forced to live days of a life not worth living, while the harm of death is being denied future days of a life that could have been worth living.[295] It certainly seems crueler to force someone to actually live a life not worth living than to deny a person days of a future life that may have been worth living. Thus, despite the fact that the length of death row reduces the harm of execution, the time on death row awaiting execution makes the overall harm far worse.

There are also inescapable ordinal proportionality problems with execution as a penalty. Suppose that the 95-year-old and the 25-year-old have the same desert basis and thus also the same desert object. The harms of their respective executions are dramatically different, since one was deprived of a few years of life and the other was deprived of decades of life.[296] This type of ordinal disproportionality also exists in Jones' case. At the time of their respective executions, Solomon was 41 years old and Jones was 72 years old. Suppose that Solomon and Jones would have otherwise each lived a total of 74 years, which is the average life span for American men as of 2021.[297] If so, then Solomon would have been deprived of 36 years of a future life, while Jones would have been deprived of two years of a future life. Thus, their respective executions harm Solomon and Jones to dramatically different degrees and are thus ordinally disproportional.

Indeed, this disproportionality is yet another reason why the pro-death-penalty retributivist should reject the second horn of the dilemma created by the appellate process. The second option for the pro-death-penalty retributivist in this dilemma is eliminating the appellate process. Though this option certainly faces the concern that undeserving people may be executed, we also see that immediate execution would still be ordinally disproportional, given that the degree of harm inflicted through execution varies dramatically from person to person. Notably, this concern applies to any jurisdiction with capital punishment, including those without appellate oversight.

---

[295] On death row, people are specifically denied agency over their own lives. If someone attempts suicide, then the person is placed on "protective" watch. If a person dies by suicide and not by state execution, it often described as "denying justice" because the punishment is being killed *by the state*, not just dying in general.
Wilkerson, "The Implications of Trying to Kill Yourself on Death Row."

[296] In the US, people are executed at drastically different ages. For instance, Walter Moody was executed at the age of 83, while Scott Carpenter was executed at the age of 22. See Blinder, "Alabama Executes Mail Bomber, 83, the Oldest Inmate Put to Death in Modern Era"; Thornton, "22-Year-Old State Man Will Be One of Youngest Executed in U.S. Age Debated as Prisoner Waives Appeals in '94 Murder."

[297] Arias et al., "Provisional Life Expectancy Estimates for 2021."

One could respond, however, that the same concern exists with incarceration, since people experience incarceration differently.[298] For instance, if a person is claustrophobic and thus experiences immense psychological distress from confinement, she may also suffer more than deserved. Thus, one could argue that the standard practice of incarceration seems to have the same concern as the one articulated regarding death. Moreover, an advocate of this response may point out that it is impossible to equalize suffering in this way. Instead, one should focus only on the state's action rather than on the individual's experience of incarceration. If the state imposes the same penalties, then people are treated proportionately, even if people suffer to different degrees from the exact same penalty. Thus, it may seem that the death penalty can use the same defense—even if people experience different levels of harm, the state treats everyone in the same way.

The solution for this problem in the context of incarceration, however, does not work in the context of execution. First, in the context of execution, the analogue of the concern that people experience incarceration differently would be the concern that people would have had different future lives. In the same way that incarceration may make a wealthy person suffer more than the average person, depriving a person of a fulfilling and happy future life is worse than depriving a person of a mediocre future life. Second, in the context of incarceration, the penalty imposed by the state is the loss of liberty for a set number of days. If two people both commit the same crime and are equally blameworthy, then the state would treat them the same way by depriving them both of their liberty for the same number of days. In the context of execution, however, the penalty is the loss of days of a life worth living, which is beyond the state's control. If two people both commit the same crime and are equally blameworthy, the state cannot deprive them both of the same number of days of a life worth living.

One could respond that the essence of the death penalty is depriving a person of life. The execution, which ends a person's life, is the primary harm. Perhaps, then, the would-be length of the person's life is not relevant to the penalty itself.

That response, however, is akin to arguing that sentence lengths are not relevant, since incarceration as a penalty is just meant to deprive a person of her liberty. If one adopts the position that only the type of penalty matters, then the length of the liberty deprivation would not be relevant. According to such a view, it would be justified to have a system in which people are locked up post-conviction and then released at different times on the basis of a random lottery system that is

[298] See, e.g., Kolber, "The Subjectivist Critique of Proportionality.".

completely non-responsive to the person's desert or crime. A retributivist committed to desert and proportionality would certainly reject such a system. Instead, in the context of incarceration, the state imposes the penalty of liberty deprivation but modulates the number of days to correspond with a person's desert. In the context of execution, however, the state imposes the penalty of depriving a person of a future life, but it cannot modulate the number of days to correspond with a person's desert. Thus, the death penalty will be inescapably disproportional in a way that incarceration is not.

Thus, the harm of execution poses a significant challenge for the pro-death-penalty retributivist. Since the harm is primarily the deprivation of days of a future life, people will be harmed by the death penalty to dramatically different extents, depending on factors that are external to their desert. Though other types of punishment face the challenge of uniformity, the death penalty is uniquely impossible to impose uniformly, since the primary harm itself is impossible to control or modulate. Thus, even execution as a penalty on its own will necessarily be disproportional. This alone is reason for the pro-death-penalty retributivist to be highly concerned about capital punishment.

A pro-death-penalty retributivist could suggest that there may be a way to standardize the days of life deprived by execution to some extent. The closest approximation would be executing both people x days before their respective natural death dates. This, however, would obviously be impossible to implement, since we do not know when a person's natural death will occur. Still, perhaps the state could approximate. Using the average age, the state could set ages for execution depending on the crime. For instance, people who commit a double homicide are executed at 65, in order to deprive them of approximately 10 years of life. People who commit a triple homicide are executed at 60, in order to deprive them of approximately 15 years of life, and so on and so forth.

Again, there are a variety of practical and theoretical problems with this approach. First, the average life span is constantly changing. Second, the average life span differs for different groups of people, depending on factors including gender, race, and socio-economic status. On the one hand, executing people at different ages depending on those protected characteristics seems highly problematic, and it would likely cause confusion for individuals about the potential penalties. On the other hand, given that average life spans do range, if those characteristics are not taken into consideration, then certainly groups that have those characteristics will be systematically over-punished (or under-punished).

The biggest problem, however, would be the incarceration that proceeds the execution. Suppose that a 25-year-old and a 65-year-old commit the exact same crime and thus have the same desert basis and desert object. Suppose further that, to properly fulfill their desert object and to treat

them equally, both should be executed at the age of 70, thus depriving both of five years of a life worth living. If so, then the 65-year-old would receive five years of incarceration on death row and then execution. The 25-year-old, however, would receive 45 years of incarceration on death row and then execution. As discussed above, long-term incarceration while awaiting execution is likely worse than death itself. As a result, the 25-year-old and the 65-year-old will be punished to dramatically different extents.

This creates a new proportionality dilemma for the pro-death-penalty retributivist. On the one hand, one could attempt to standardize the length of time on death row. If so, then one faces the disproportionality inherent to executing people at different ages. On the other hand, one could attempt to standardize the age of execution to ensure that people are deprived of roughly the same number of years of life. If so, then one faces the disproportionality of varied lengths of long-term incarceration pre-execution. Either way, the punishments will be ordinally disproportional.

Thus, there is no way to standardize the harms inflicted through capital punishment—any attempt to standardize one type of harm will necessarily mean that the other type of harm cannot be standardized. Thus, the pro-death-penalty retributivist faces an inescapable disproportionality dilemma that is inherent to the practice of capital punishment.

Conclusion

In this paper, I set forward a novel proportionality challenge for the pro-death-penalty retributivist. I started by setting up the tradeoff created by an appellate process: the post-conviction review is needed in order to ensure that only the deserving are executed, but the incarceration while the appellate process runs its course means that people may be harmed more than they deserve. Moreover, given that the length of the appellate process is not contingent on desert-related factors, ordinal disproportionality is bound to occur.

The pro-death-penalty retributivist faces her first dilemma: either maintain the current appellate process with all of its disproportionality or abolish the appellate process and risk executing the undeserving. Any attempt to standardize the length of the appellate process, however, leads the pro-death-retributivist into a new dilemma. If one standardizes the length of the appellate process, then people will be disproportionately harmed by execution, since the degree of harm is contingent on factors external to desert, such as age. But, if one standardizes the degree of harm imposed by execution (such as through an age-based sentencing scheme), then people will be disproportionality harmed by the long-term incarceration awaiting their execution age.

Many anti-death penalty arguments in the literature attack the premises that pro-death-penalty retributivists rely upon to justify the death penalty. My argument, however, completely adopts all of the premises of the pro-death-penalty retributivist. Even with these commitments, I point out that the harms involved with capital punishment lead to inevitable disproportionality. In doing so, I demonstrate that the death penalty, a punishment that is often cited as the clearest example of proportional punishment, actually faces inescapable proportionality problems.

Furthermore, the concerns raised in this paper are unique to capital punishment. Standard forms of punishment, like years of incarceration, do not have the same issues. First, death is final and irreparable, while incarceration is not. Given the finality of death, the need to be certain that the person deserves death increases. Second, the anticipation of execution is its own extreme harm that does not exist with standard sentences. Even with life sentences, people do not experience the psychological torture of awaiting execution by the state. Third, the time spent incarcerated while the appellate process runs its course is determined by non-desert factors, while the time spent incarcerated serving a sentence is determined by desert. Finally, the degree of harm imposed by execution is out of the state's control. While the state can control the length of a person's sentence, the state cannot control, for instance, the age at which a person commits a crime.

Thus, it seems that the best way out of the dilemma for the pro-death-penalty retributivist would be to abolish the death penalty and replace it with a punishment, like incarceration, that does not have these temporal proportionality problems.

Bibliography

Alexander, Larry. "Proportionality's Function." *Criminal Law and Philosophy* 15, no. 3 (October 1, 2021): 361–72. https://doi.org/10.1007/s11572-021-09564-x.

———. "Retributivism and the Inadvertent Punishment of the Innocent." *Law and Philosophy* 2, no. 2 (August 1, 1983): 233–46. https://doi.org/10.1007/BF00144449.

Alexander, Larry, and Kimberly Kessler Ferzan. *Crime and Culpability: A Theory of Criminal Law*. Cambridge Introductions to Philosophy and Law. Cambridge: Cambridge University Press, 2009. https://doi.org/10.1017/CBO9780511804595.

Arias, Elizabeth, Betzaida Tejada-Vera, Kenneth D. Kochanek, and Farida B. Ahmad. "Provisional Life Expectancy Estimates for 2021." Vital Statistics Rapid Release. National Center for Health Statistics, August 2022.

Atkins v. Virginia, 536 U.S. 304 (2002) (June 20, 2002).

Bagaric, Mirko, and James McConvill. "Giving Content to the Principle of Proportionality: Happiness and Pain as the Universal Currency for Matching Offence Seriousness and Penalty Severity." *The Journal of Criminal Law* 69, no. 1 (2005): 50–74. https://doi.org/10.1350/jcla.69.1.50.58954.

Bedau, Hugo Adam. "The Minimal Invasion Argument Against the Death Penalty." *Criminal Justice Ethics* 21, no. 2 (June 1, 2002): 3–8. https://doi.org/10.1080/0731129X.2002.9992122.

Berns, Walter. "For Capital Punishment." *Harper's; New York, N.Y.*, 1979.
Blinder, Alan. "Alabama Executes Mail Bomber, 83, the Oldest Inmate Put to Death in Modern Era." *The New York Times*, April 20, 2018, sec. U.S. https://www.nytimes.com/2018/04/19/us/alabama-execution-walter-leroy-moody.html.
Blume, John H. "Killing the Willing: Volunteers, Suicide and Competency." *Michigan Law Review* 103 (2005): 939.
Bonta, James, and Paul Gendreau. "Reexamining the Cruel and Unusual Punishment of Prison Life." *Law and Human Behavior* 14 (1990): 347–72. https://doi.org/10.1007/BF01068161.
Boonin, David. "The Appeal to Necessity." In *The Problem of Punishment*. Cambridge University Press, 2008.
Bowcott, Owen, and Maya Wolfe-Robinson. "British Court to Rule on Death Sentences for Two Trinidad Murderers." *The Guardian*, February 4, 2015, sec. Law. https://www.theguardian.com/law/2015/feb/04/british-court-to-rule-on-death-sentences-for-two-trinidad-murderers.
Brooks, Thom. "Retribution and Capital Punishment." In *Retributivism*. Oxford University Press, 2011. https://oxford.universitypressscholarship.com/view/10.1093/acprof:oso/9780199752232.001.0001/acprof-9780199752232-chapter-012.
———. "Retributivist Arguments against Capital Punishment." *Journal of Social Philosophy* 35, no. 2 (2004): 188–97. https://doi.org/10.1111/j.1467-9833.2004.00224.x.
Camus, Albert. "Reflections on the Guillotine." In *Resistance, Rebellion, and Death*, translated by Justin O'Brien, 131–79. London: H. Hamilton, 1961.
Caruso, Gregg D. "Free Will Skepticism and Criminal Behavior: A Public Health-Quarantine Model." *Southwest Philosophy Review* 32, no. 1 (August 21, 2016): 25–48. https://doi.org/10.5840/swphilreview20163214.
———. "Justice Without Retribution: An Epistemic Argument Against Retributive Criminal Punishment." *Neuroethics* 13, no. 1 (2020): 13–28. https://doi.org/10.1007/s12152-018-9357-8.
Chammah, Maurice, and Keri Blakinger. "They Are Terminally Ill. States Want to Execute Them Anyway." The Marshall Project, May 6, 2021. https://www.themarshallproject.org/2021/05/06/they-are-terminally-ill-states-want-to-execute-them-anyway.
Christopher, Russell. "Deterring Retributivism: The Injustice of 'Just' Punishment." *Northwestern University Law Review* 96 (2002): 843–976.
Coleman, James E. Jr. "Litigating at the Speed of Light: Postconviction Proceedings under a Death Warrant." *Litigation* 16, no. 4 (1990 1989): 14–53.
Cottingham, John. "Varieties of Retribution." *Philosophical Quarterly* 29 (1979): 238–46.
Davis, Michael. "A Sound Retributive Argument for the Death Penalty." *Criminal Justice Ethics* 21, no. 2 (2002): 22–26. https://doi.org/10.1080/0731129X.2002.9992125.
———. "Harm and Retribution." *Philosophy & Public Affairs* 15, no. 3 (1986): 236–66.
———. "How to Make the Punishment Fit the Crime." *Ethics* 93, no. 4 (1983): 726–52.
Death Penalty Information Center. "DPIC Adds Eleven Cases to Innocence List, Bringing National Death-Row Exoneration Total to 185." *Deathpenaltyinfo.Org*, February 18, 2021. https://deathpenaltyinfo.org/news/dpic-adds-eleven-cases-to-innocence-list-bringing-national-death-row-exoneration-total-to-185.
"Death Sentences and Executions 2018." Amnesty International, 2018.
Dolinko, David. "How to Criticize the Death Penalty." *The Journal of Criminal Law and Criminology (1973-)* 77, no. 3 (1986): 546–601. https://doi.org/10.2307/1143430.
Duff, R. A. *Punishment, Communication, and Community*. Oxford University Press, 2000.

Finkelstein, Claire. "Death and Retribution." *Criminal Justice Ethics* 21, no. 2 (June 2002): 12–21. https://doi.org/10.1080/0731129X.2002.9992124.

Flanders, Chad. "Time, Death, and Retribution." *University of Pennsylvania Journal of Constitutional Law* 19 (2017): 431.

Fogel, Alan. *Developing through Relationships: Origins of Communication, Self, and Culture*. Developing through Relationships: Origins of Communication, Self, and Culture. Chicago, IL, US: University of Chicago Press, 1993.

Fournier, Ron. "The Time Bill Clinton and I Killed a Man." *The Atlantic*, May 28, 2015. https://www.theatlantic.com/politics/archive/2015/05/the-time-bill-clinton-and-i-killed-a-man/460869/.

Gardner, John. "Crime: In Proportion and in Perspective." In *Offences and Defences: Selected Essays in the Philosophy of Criminal Law*. Oxford University Press, 2007.

Gerstein, Robert S. "Capital Punishment-"Cruel and Unusual"?: A Retributivist Response." *Ethics* 85, no. 1 (October 1974): 75–79. https://doi.org/10.1086/291941.

Haag, Ernest van den. "The Ultimate Punishment: A Defense." *Harvard Law Review* 99, no. 7 (1986): 1662. https://doi.org/10.2307/1341082.

Halvorsen, Vidar. "Is It Better That Ten Guilty Persons Go Free than That One Innocent Person Be Convicted." *Criminal Justice Ethics* 23 (2004): 3.

Hart, H. L. A. *Punishment and Responsibility: Essays in the Philosophy of Law*. Oxford: Clarendon P., 1968.

Hart, H.L.A. "Postscript: Responsibility and Retribution." In *Punishment and Responsibility*. Oxford University Press, 1968.

Hirsch, Andrew von. "Proportionality in the Philosophy of Punishment." *Crime and Justice* 16 (1992): 55–98.

Hsu, Jimmy Chia-Shin. "Does Communicative Retributivism Necessarily Negate Capital Punishment?" *Criminal Law and Philosophy* 9, no. 4 (2015): 603–17. https://doi.org/10.1007/s11572-013-9261-6.

Husak, Douglas. "Already Punished Enough." *Philosophical Topics* 18, no. 1 (1990): 79–99.

Johnson, Sheri Lynn, John H. Blume, and Brendan Van Winkle. "Atkins v. Virginia at Twenty: Still Adaptive Deficits, Still in the Developmental Period." SSRN Scholarly Paper. Rochester, NY, August 18, 2022. https://papers.ssrn.com/abstract=4194005.

Kadish, Sanford. "The Criminal Law and the Luck of the Draw." *Journal of Criminal Law and Criminology* 84, no. 4 (1994): 25.

Kershnar, Stephen. "A Defense of Retributivism." *International Journal of Applied Philosophy* 14, no. 1 (April 1, 2000): 97–117. https://doi.org/10.5840/ijap20001416.

Kleinig, John. *Punishment and Desert*. The Hague: Martinus Nijhoff, 1973.

Kolber, Adam J. "Against Proportional Punishment." *Vanderbilt Law Review* 66, no. 4 (2013): 1141–79.

———. "The Subjectivist Critique of Proportionality." In *The Palgrave Handbook of Applied Ethics and the Criminal Law*, edited by Larry Alexander and Kimberly Kessler Ferzan, 571–95. Cham: Springer International Publishing, 2019. https://doi.org/10.1007/978-3-030-22811-8_24.

Kovarsky, Lee. "The American Execution Queue." *Stanford Law Review* 71 (2019): 1163–1228.

Lempert, Richard O. "Desert and Deterrence: An Assessment of the Moral Bases of the Case for Capital Punishment." *Michigan Law Review* 79, no. 6 (May 1981): 1177. https://doi.org/10.2307/1288113.

Lenta, Patrick, and Douglas Farland. "Desert, Justice and Capital Punishment." *Criminal Law and Philosophy* 2, no. 3 (October 1, 2008): 273–90. https://doi.org/10.1007/s11572-008-9045-6.

Lippert-Rasmussen, Kasper. "Why Killing Some People Is More Seriously Wrong than Killing Others." *Ethics* 117, no. 4 (July 2007): 716–38. https://doi.org/10.1086/519064.

McCorkle, Richard. "Personal Precautions to Violence." *Criminal Justice and Behavior* 19, no. 2 (1992): 160–73. https://doi.org/10.1177/0093854892019002004.

McDermott, Daniel. "A Retributivist Argument against Capital Punishment." *Journal of Social Philosophy* 32 (2001): 317–33.

McMahan, Jeff. "Causing People to Exist and Saving People's Lives." *The Journal of Ethics* 17, no. 1 (June 1, 2013): 5–35. https://doi.org/10.1007/s10892-012-9139-1.

———. "Death and the Value of Life." *Ethics* 99, no. 1 (October 1988): 32–61. https://doi.org/10.1086/293034.

Medley, Petitioner, 134 U.S. 160 (1890) (March 3, 1890).

Millum, Joseph. "Age and Death: A Defence of Gradualism." *Utilitas* 27, no. 3 (September 2015): 279–97. https://doi.org/10.1017/S0953820815000047.

Nagel, Thomas. "Death." *Noûs* 4, no. 1 (1970): 73–80. https://doi.org/10.2307/2214297.

People v Anderson, No. 493 P 2d 880 (California Supreme Court 1972).

Pojman, Louis P., and Owen McLeod. *What Do We Deserve?: A Reader on Justice and Desert*. Oxford: Oxford University Press, 1998.

ProCon. "Death Row Inmates." Britanica ProCon.org. Accessed February 5, 2023. https://deathpenalty.procon.org/death-row-inmates/.

Rehman, Javine, Priya Gopalan, Matthew Gillet, Ganna Yudkivska, Miriam Estrada-Castillo, and Mumba Malila. "Iran: UN Experts Condemn Recent Executions, Urge Moratorium on Death Penalty." Press Release. United Nations Human Rights Office of the High Commissioner, May 9, 2023. https://www.ohchr.org/en/press-releases/2023/05/iran-un-experts-condemn-recent-executions-urge-moratorium-death-penalty.

Roark, Anthony P. "Retribution, the Death Penalty, and the Limits of Human Judgment." *International Journal of Applied Philosophy* 13, no. 1 (April 1, 1999): 57–68. https://doi.org/10.5840/ijap19991317.

Roberts-Cady, Sarah. "Against Retributive Justifications of the Death Penalty." *Journal of Social Philosophy* 41, no. 2 (2010): 185–93. https://doi.org/10.1111/j.1467-9833.2010.01486.x.

Segura, Liliana. "Facing Execution at 72, Georgia's Oldest Death Row Inmate Exposes Death Penalty's Racist Roots." *The Intercept*, January 31, 2016. https://theintercept.com/2016/01/31/brandon-astor-jones-georgia-death-row-inmate-reminder-of-death-penalty-racist-roots/.

Shafer-Landau, Russ. "Retributivism and Desert." *Pacific Philosophical Quarterly* 81, no. 2 (June 1, 2000): 189–214. https://doi.org/10.1111/1468-0114.00102.

Sher, George. *Desert*. Princeton University Press, 1987.

Sigler, Mary. "Private Prisons, Public Functions, and the Meaning Punishment." *Florida State University Law Review* 38, no. 1 (2010).

Slifer, Stephanie. "How the Wrongfully Convicted Are Compensated for Years Lost." CBS News, March 27, 2014. https://www.cbsnews.com/news/how-the-wrongfully-convicted-are-compensated/.

Smith, Amy. "Not 'Waiving' But Drowning: The Anatomy of Death Row Syndrome and Volunteering for Execution." *The Boston University Public Interest Law Journal* 17, no. 2 (May 21, 2008): 237–54.

Snell, Tracy L. "Capital Punishment, 2019 - Statistical Tables." U.S. Department of Justice, Bureau of Justice Statistics, 2019.

Sorell, Tom. "Aggravated Murder and Capital Punishment." *Journal of Applied Philosophy* 10, no. 2 (1993): 201–13. https://doi.org/10.1111/j.1468-5930.1993.tb00076.x.

Steiker, Carol S., and Jordan M. Steiker. "A Tale of Two Nations: Implementation of the Death Penalty in Executing versus Symbolic States in the United States Symposium: Punishment Law and Policy: II. Death Penalty in Practice." *Texas Law Review* 84, no. 7 (2006 2005): 1869–1928.

Stevenson, Bryan A. "The Politics of Fear and Death: Successive Problems in Capital Federal Habeas Corpus Cases." *New York University Law Review* 77 (2002): 699–795.

Tabachnick, Cara. "Ohio Governor Postpones 3 Executions amid Struggle to Find Lethal Drug Suppliers." *CBS News*, April 15, 2023. https://www.cbsnews.com/news/ohio-governor-mike-dewine-postpones-three-executions-lethal-drug-suppliers/.

Tartaro, Christine, and David Lester. "Suicide on Death Row." *Journal of Forensic Sciences* 61, no. 6 (2016): 1656–59. https://doi.org/10.1111/1556-4029.13069.

Thornton, Anthony. "22-Year-Old State Man Will Be One of Youngest Executed in U.S. Age Debated as Prisoner Waives Appeals in '94 Murder." *The Oklahoman*, April 8, 1997. https://www.oklahoman.com/story/news/1997/04/08/year-old-state-man-will-one-youngest-executed-us-age-debated-as-prisoner-waives-appeals-in-94-murder/62318508007/.

Toch, Hans. *Men in Crisis: Human Breakdowns in Prison*. Transaction Publishers, 2007.

Totenberg, Nina, and Domenico Montanaro. "Supreme Court Closely Divides On 'Cruel and Unusual' Death Penalty Case." *NPR*, April 1, 2019, sec. Law. https://www.npr.org/2019/04/01/708729884/supreme-court-rules-against-death-row-inmate-who-appealed-execution.

United States v. Salerno 481 U.S. 739 (1987) (May 26, 1987).

Wilkerson, George. "The Implications of Trying to Kill Yourself on Death Row." *The Marshall Project*, January 26, 2017, sec. Life Inside. https://www.themarshallproject.org/2017/01/26/the-implications-of-trying-to-kill-yourself-on-death-row.

Wise, Alana. "The Justice Department Is Pausing Federal Executions After They Resumed Under Trump." *NPR*, July 1, 2021, sec. Politics. https://www.npr.org/2021/07/01/1012366520/the-justice-department-is-pausing-federal-executions-after-they-resumed-under-tr.

Yaffe, Gideon. "The Norm Shift Theory of Punishment." *Ethics* 132, no. 2 (January 2022): 478–507. https://doi.org/10.1086/716874.

Yost, Benjamin S. *Against Capital Punishment*. New York: Oxford University Press, 2019.

———. "What's Wrong with Differential Punishment?" *Utilitas* 29, no. 3 (September 2017): 257–85. https://doi.org/10.1017/S095382081600039X.

Compensating the Guilty over the Innocent

The view that people who are exonerated are owed compensation is fairly uncontroversial. But is compensation owed when a guilty person is over-punished? Drawing primarily on the literature on the exculpatory power of false beliefs, I demonstrate that there is a stronger argument for compensation when the state's action is based on a false moral belief than when the state's action is based on a false factual belief. I then show that incorrectly punishing an innocent person is typically a wrong action based on a false factual belief, while over-punishing a guilty person is typically a wrong action based on a false moral belief. Through five uncontroversial premises, I arrive at the surprising conclusion that compensation for over-punishing a guilty person is more defensible than compensation for incorrectly punishing an innocent person.

Introduction

On July 3, 2020, Anthony Mazza was released from prison after serving 47 years for a crime he did not commit. Mazza had been sentenced to life without parole in 1972 after being framed for a robbery-murder by the actual perpetrator. At the time of his release, Mazza was 73 years old. Of the over 3,000 people exonerated in the US since 1989, Mazza spent the longest time incarcerated.[299]

Let's compare Mazza's story with that of Leandro Andrade. In November 1995, Andrade—a father of three—was caught shoplifting $153 worth of children's videos from a K-Mart in California. By doing so, Andrade hit his "third strike," meaning that he had previously been convicted of two other crimes. Though all of the crimes had been non-violent, minor thefts, Andrade was sentenced to life in prison, without the possibility of parole for 50 years.[300] Andrade and hundreds of others with similar stories were sent to prison for life under California's three strikes law that imposed a life sentence for a person's third offense.[301] Unlike Mazza, Andrade did commit a crime. Yet, a life sentence for shoplifting seems unjust.

Do Mazza and Andrade deserve compensation for these seemingly undeserved long-term periods of incarceration? When it comes to people like Mazza who are completely innocent, most would agree that the state owes compensation. Indeed, many jurisdictions—including the United

[299] Otterbourg, "National Registry of Exonerations: Anthony Mazza." All future references to this case are from here.
[300] Chemerinsky, "Cruel and Unusual." All future references to this case come from this source.
[301] Jaffe, "Cases Show Disparity of California's 3 Strikes Law."

States, England and Wales, and most European countries—offer compensation to people who are exonerated.[302] The requirement to compensate exonerees is even enshrined in international law.[303] When it comes to people like Andrade, who have committed a crime but appear to be over-punished, the question of compensation is far less settled. It has received very little philosophical attention, and there appear to be no legal requirements to provide compensation in such cases.[304]

In this paper, I argue for the surprising conclusion that compensation in cases like Andrade's is actually more defensible than compensation in cases like Mazza's. This is not to say that innocent people like Mazza who spend years incarcerated should not be compensated. Rather, my argument is meant to demonstrate that if one supports compensation for innocent people who are exonerated, then *a fortiori* one should support compensation for guilty people who are over-punished. Thus, if we think that the practice of compensating exonerees is justified, then we should certainly support compensating those who are correctly convicted but over-punished.

The argument proceeds as follows: In §1, I provide the definitions of incorrect punishment (punishing the innocent) and over-punishment (over-punishing the guilty). In §§2-3, I raise the question of compensation for over-punishment and present the central argument. In §4, I argue that a wrong action based on a blameless false factual belief is a blameless mistake, while a wrong action based on a blameworthy false moral belief is a blameworthy harm.[305] In §5, I demonstrate that a typical case of punishing an innocent person is a wrong action based on a blameless false factual belief, while a typical case of over-punishing a guilty person is a wrong action based on a blameworthy false moral belief. In §6, I contend that the argument for compensation for blameworthy harms is stronger than the argument for compensation for blameless mistakes. Thus, the argument for compensation for over-punishing the guilty is stronger than the argument for incorrectly punishing the innocent. In §7, I apply the framework to non-standard cases of incorrect punishment and over-punishment. Finally,

---

[302] For a 2023 comparative study of compensation statutes for incorrect punishment in Europe and the US, see: Jasiński and Kremens, *Compensation for Wrongful Convictions*. For an overview of the US specifically, see: Ryan, "Compensation for Wrongful Convictions in the United States." Unfortunately, it is often complicated for exonerated people to receive compensation despite statutes that mandate compensation. According to the National Registry of Exonerations, only 41% of exonerees who have sought compensation have in fact been compensated in the United States. Janse and Shapiro, "For the Exonerated, Compensation Is a Battle for Stability and Dignity."

[303] International Covenant on Civil and Political Rights (ICCPR), Article 14.

[304] To my knowledge, no such statute exists that guarantees compensation in such cases. The statutes regarding compensation for exonerees are not accompanied by statutes regarding compensation for over-punishment. For instance, the International Covenant on Civil and Political Rights (ICCPR) makes no mention of over-punishment.

[305] In the literature, some people use the term "factual belief," and others use the term "non-moral belief," depending on a given scholar's meta-ethical commitments. For the purposes of this paper, those two terms could be used interchangeably, and I have opted to only use the term "factual belief" for clarity and consistency.

in §8, I consider how the arguments apply to a legitimate state, given its unique coercive powers and duties.

§1. Definitions

Before we begin, it is helpful to clarify the relevant terms. When an innocent person is mistakenly convicted, she is then punished, typically through incarceration. I will refer to the harm experienced by the innocent person as **incorrect punishment**.

When a guilty person is correctly convicted, the state imposes a punishment, which is also typically incarceration. Intuitively, there are limits to punishing people, even if they have committed a crime. There are a variety of harms that may be inflicted on people who have committed crimes that could fall beyond the limits of justified punishment. Certain types of punishment, like the death penalty, could be harms that fall beyond the limits of justified punishment, if capital punishment is never permissible to impose.[306] For those who support the death penalty, other types of punishment, like torture, rape, and bodily mutilation, may fall beyond the limits of justified punishment. Alternatively, harms that extend beyond the length of justified punishment—either before proper conviction or after a completed sentence—could similarly be instances of harms that fall beyond the limits of justified punishment, if punishment is time-bound. For instance, much of Andrade's sentence would likely fall beyond the limits of justified punishment.

For the purposes of this paper, I remain agnostic as to the content of the limits of justified punishment. If there are limits to justified punishment, then there similarly must be harms that can occur beyond those limits. I will refer to the harm beyond justified punishment experienced by someone who has committed a crime as **over-punishment**.

Though over-punishment encompasses all cases of harms beyond the justified limits, I will set aside instances of relatively minor over-punishment for the purposes of this paper. Even just one extra day of incarceration would qualify as over-punishment according to my definition. For this paper, however, I primarily consider harms that are clearly and significantly beyond the limits of justified punishment.

[306] See, e.g., Bedau, "The Minimal Invasion Argument Against the Death Penalty"; Nathanson, *An Eye for an Eye*; Roberts-Cady, "Against Retributive Justifications of the Death Penalty"; Yost, "Death and Retribution"; Brooks, "Retributivist Arguments against Capital Punishment."

§2. Central Question

Compensation for incorrect punishment is widely legislated and widely supported. Compensation for incorrect punishment is even required by international law. Article 14 of the International Covenant on Civil and Political Rights (ICCPR) notes a requirement for compensation in cases in which a person is wrongfully convicted:

> When a person has by a final decision been convicted of a criminal offence and when subsequently his conviction has been reversed or he has been pardoned on the ground that a new or newly discovered fact shows conclusively that there has been a miscarriage of justice, *the person who has suffered punishment as a result of such conviction shall be compensated according to law*, unless it is proved that the nondisclosure of the unknown fact in time is wholly or partly attributable to him.[307]

According to a comparative international law study, many, if not all, countries in Europe have statutes that require compensation for incorrect punishment.[308] In England and Wales, compensation is owed in cases of "miscarriages of justice," which most clearly applies in instances of incorrect punishment.[309] In the US, the federal government guarantees $50,000 per year of incorrect punishment, with an additional $50,000 for every year spent on death row.[310] Thus, there is a common legal requirement for the state to compensate individuals who are incorrectly punished.

Furthermore, there is overwhelming support for compensating innocent people who are incorrectly punished in the literature. Many scholars take the requirement for compensation for incorrect punishment as a "necessary condition to legitimatize the criminal justice system."[311] The primary question in the literature is not if people who are wrongly convicted should receive compensation. Rather, scholars focus on whether exonerees are sufficiently compensated, how easy it is for exonerees to receive compensation, and whether non-monetary support is owed to exonerees.[312]

---

[307] International Covenant on Civil and Political Rights (ICCPR), Article 14 (emphasis mine). For other mentions of incorrect punishment in international law, see Article 3 of Protocol No. 7 of the European Convention on Human Rights (ECHR), further interpreted by the European Court of Human Rights (ECtHR).
[308] Jasiński and Kremens, *Compensation for Wrongful Convictions*.
[309] Quirk, "Compensation for Wrongful Convictions in England and Wales."
[310] Slifer, "How the Wrongfully Convicted Are Compensated for Years Lost."
[311] See, e.g., "Jasiński and Kremens, *Compensation for Wrongful Convictions*, 2.
[312] For instance, Jasiński and Kremens argue that the "crucial question is how the compensation system…should operate" (Jasiński and Kremens, 2). More specifically, for arguments about how compensation for wrongful conviction statutes are too restrictive, see, e.g., Kahn, "Presumed Guilty until Proven Innocent"; Kaiser, "Wrongful Conviction and Imprisonment"; Madrigal and Norris, "The Good, the Bad, and the Uncertain"; Faridi, Hoffman, and Montuori, "Undoing Time"; Zalman and Carrano, *Wrongful Conviction and Criminal Justice Reform*. For arguments that compensation and more is owed, see, e.g., Chunias and Aufgang, "Beyond Monetary Compensation"; Armbrust, "When Money Isn't Enough"; Simms, "Statutory Compensation for the Wrongly Imprisoned."

Moreover, numerous psychological studies have demonstrated that many people support compensation, often more than what is legally owed, for people who are exonerated.[313] Given the widespread legal requirement, as well as the public and scholarly support for the practice, we can conclude that compensation for incorrect punishment is largely straightforward and noncontroversial.

But what should happen in cases in which guilty people are over-punished? To my knowledge, this topic has received little to no scholarly attention. The statutes requiring compensation for incorrect punishment are not accompanied by statutes requiring compensation for over-punishment.[314]

This discrepancy may seem unsurprising. When the state incorrectly punishes a person, it takes an innocent person off the street and locks her away for years, if not decades, of her life. Though the years of her life cannot be returned, monetary compensation seems like the next best option.

In the context of over-punishment, however, compensation is less obviously due. Intuitively, the idea of punishing a completely innocent person seems far worse than over-punishing a guilty person. In the coming sections, I argue for the exact opposite: there is a stronger case for compensation in cases of over-punishment than in cases of incorrect punishment.

§3. Compensation for Over-Punishment

First, I argue that incorrect punishment and over-punishment are the same type of harm. In both cases, the state uses the criminal legal system to harm a person in a way that is unjustified. Thus, the state is using its coercive force—through deprivations of liberty, bodily autonomy, etc.—in ways that are impermissible, thus violating the rights of the incorrectly punished person or the over-punished person in the same way. In this view, I follow Chris Wellman, a prominent punishment theorist, who notes:

> In my view, those who are over-punished have their rights violated in the same way and for the same reasons as the innocent who are wrongly punished. If a criminal has forfeited her right against a one-year incarceration but is imprisoned for two years, for instance, then the

[313] See, e.g., Nowotny, Shlosberg, and McAndrew, "Understanding Public Views of Wrongful Conviction Frequency and Government Responsibility for Compensation"; Karaffa, Page, and Koch, "Compensating the Innocent: Perceptions of Exonerees' Deservingness to Receive Financial Compensation for Wrongful Convictions"; Clow et al., "Public Perception of Wrongful Conviction."

[314] See, e.g., the absence of a mention of over-punishment in the ICCPR, which only states that compensation is owed for incorrect punishment. The US Code even stipulates that federal compensation is only owed when the person is found "not guilty" (which is meant to exclude cases in which a person is pardoned). See 28 U.S. Code § 2513 - Unjust Conviction and Imprisonment; 28 U.S. Code § 1495 - Damages for unjust conviction and imprisonment; claim against United States.

> second year in prison would be an impermissible over-punishment… [T]his unjustified second year of incarceration for the guilty person is just as objectionable as a one-year prison sentence for an innocent person.[315]

Thus, the way in which a person's interests are set back in cases of incorrect punishment and over-punishment are quite similar, if not identical.

One may object that there may be different types of harms involved in incorrect punishment, such as reputational or emotional harms. I argue that the distinctions between types of harms are often not categorical distinctions, but rather track degrees of harm. The fact that in one case the setback comes from physical pain and in another case the setback comes from reputational damage is immaterial. Rather, all harm is a setback of one's interest. It may be the case that reputational harms often set back one's interests less than physical harms. Yet, there will also be cases in which a reputational harm sets back one's interests much more than a physical harm. Furthermore, when applied to cases of incorrect punishment and over-punishment, a person's interests may be set back in similar ways. In both instances, people lose their liberty, potentially for many years of their lives. Thus, not only do incorrect punishment and over-punishment both set back one's interest, the two set back one's interests in similar ways.

One may reply, however, that there are necessarily different degrees of harm involved in incorrect punishment and over-punishment. Perhaps incorrect punishment is always more harmful than over-punishment. If so, then the two may be disanalogous in terms of the extent of the harm.

I have two responses. First, while there may be certain cases of incorrect punishment which are worse than certain cases of over-punishment, it is not the case that incorrect punishment is categorically worse. Compare the following two examples: An innocent person is wrongly convicted and spends one week incarcerated. A guilty person is correctly convicted, and while she deserves only one month of incarceration, she receives 30 years of incarceration. In this comparison, the case of over-punishment certainly seems more harmful than the case of incorrect punishment.

Second, though degrees of harm may be difficult to compare, both incorrect punishment and over-punishment come in degrees. It is often difficult to compare degrees of harm when people experience different treatments. For instance, it is not clear which is worse: a day of torture or 10 years of incarceration. Thus, it may be hard to determine the comparative degree of harm between two cases. But there will also be plenty of instances in which it is relatively straightforward to compare a

[315] Wellman, "No One Defends the Status Quo," 180.

case of incorrect punishment and a case of over-punishment. Per Wellman's example, if an innocent person is incarcerated for one year and a guilty person is incarcerated for one year more than she deserves, it seems that the degrees of harm are fairly similar.

Given the similarities between the harms, I arrive at my central argument, which I term the **Compensation for Over-Punishment** argument. The argument is as follows:

1. A wrong action based on a blameless false factual belief is a blameless mistake.
2. A wrong action based on a blameworthy false moral belief is a blameworthy harm.
3. A typical case of incorrect punishment is a wrong action based on a blameless false factual belief.
4. A typical case of over-punishment is a wrong action based on a blameworthy false moral belief.

SC 1: A typical case of incorrect punishment is a blameless mistake. (1, 3)

SC 2: A typical case of over-punishment is a blameworthy harm. (2, 4)

5. The argument for compensation for a blameworthy harm is stronger than the argument for compensation for a blameless mistake.

Conclusion: The argument for compensation for a typical case of over-punishment is stronger than the argument for compensation for a typical case of incorrect punishment. (SC1, SC2, 5)

Thus, this argument leads us to the **asymmetry conclusion**. Since there is a stronger argument for compensation in cases of over-punishment, there is an asymmetry between incorrect punishment and over-punishment.

In the coming sections, I will defend the five assumptions in the above argument. If the five assumptions are indeed correct, then we land at our surprising conclusion that there is a stronger reason to provide compensation for a typical case of over-punishment than for a typical case of incorrect punishment.

§4. Actions on the Basis of False Beliefs

First, I defend the following two premises:

1. A wrong action based on a blameless false factual belief is a blameless mistake.
2. A wrong action based on blameworthy false moral beliefs is a blameworthy harm.

To defend these premises, let us turn to the philosophical literature about the exculpatory power of false beliefs.

4.1 False Factual Beliefs

Let us start with false factual beliefs. Generally, when a person commits a wrong action that harms another person, that action is a blameworthy harm. Some instances of wrong actions may be excused, such as when the action is performed by a young child or by a sleepwalking adult. One circumstance in which a person's wrong action may be excused is when the wrong action is based on a false factual belief. Consider the following case presented by Elizabeth Harman:

> **Cyanide**: Anne spoons cyanide into Bill's coffee, but she does not know it is cyanide: she believes that she is spooning sugar.[316]

In this case, Anne is not blameworthy for poisoning Bill if she is not blameworthy for her false factual belief regarding the sugar. Since Anne had no reason to even consider the possibility that the sugar had been swapped out for cyanide, she is not blameworthy for her false factual belief. Since Anne is not blameworthy for her false factual belief, Anne is not blameworthy for accidentally poisoning Bill.

Importantly, a reckless false factual belief does not exculpate. In **Cyanide**, Anne's belief that she is spooning sugar into Bill's coffee is epistemically reasonable and responsible. In contrast, consider another version of **Cyanide**:

> **Cyanide & Sugar**: Anne knows that she has both a container of cyanide and a container of sugar in her kitchen. She knows that the two containers can be easily confused for one another. She believes that the container in front of her is sugar, but she does not bother to check. Anne thinks she is putting sugar into Bill's coffee, but she is actually putting cyanide into Bill's coffee.

In **Cyanide & Sugar**, Anne's action seems blameworthy. Though Anne possesses a false factual belief that the substance in front of her is sugar, that false factual belief is a reckless and epistemically irresponsible belief. Given the stakes at hand, Anne should have checked the substance in front of her to confirm that it was sugar. Her reckless false factual belief does not exculpate. Rather, a false factual belief is exculpatory only if the false factual belief itself is blameless. In **Cyanide**, then, Anne's action is a blameless mistake even though Bill is harmed.

A further clarification is needed: a false factual belief does not guarantee that the resulting action is a blameless mistake. Rather, a false factual belief is only exculpatory when an agent's other relevant moral beliefs are correct. In **Cyanide**, all of Anne's relevant moral beliefs are correct; she knows that it is impermissible to poison Bill. She ultimately poisons Bill only because of a false factual belief. In contrast, consider this modified version of **Cyanide**:

---

[316] Harman, "Does Moral Ignorance Exculpate?," 444.

Cyanide & Ricin: Anne wants to poison Bill. She has two poisons in front of her, one is cyanide, and one is ricin. Anne thinks she is spooning ricin into Bill's coffee, but she is actually spooning cyanide into Bill's coffee.

In **Cyanide & Ricin**, Anne's action is certainly blameworthy. Anne possesses a false moral belief that it is permissible to poison Bill. Anne also possesses a false factual belief that she is putting cyanide into Bill's coffee. The presence of the false factual belief does not exculpate. Rather, in order for a false factual belief to be exculpatory, the agent's other relevant moral beliefs need to be correct.

To account for these clarifications, Premise 1 should be refined as follows:

1.* A wrong action based on a blameless false factual belief is a blameless mistake if and only if the action would have been permissible had the false factual belief been true.

In order to account for 1*, Premise 3 would need to be modified as follows:

3.* A typical case of incorrect punishment is a wrong action based on a blameless false factual belief where the action would have been permissible had the false factual belief been true.

Now that we have refined the premises, let us turn to the question of false moral beliefs.

4.2 False Moral Beliefs

As shown above, a false factual belief can excuse a person's wrong action. The question, then, is whether false moral beliefs operate the same way. Broadly speaking, there are two opposing responses to this question, which are best articulated by Elizabeth Harman and Gideon Rosen.[317]

Rosen argues that false factual beliefs and false moral beliefs exculpate under the same conditions. Ignorance—both of factual facts and of moral facts—can be exculpatory if the ignorance is blameless.

Rosen's View: Whenever an agent acts from ignorance, whether factual or moral, he is culpable for the act only if he is culpable for the ignorance from which he acts.[318]

The view applies to both false moral beliefs and false factual beliefs. Just as one is not blameworthy for possessing a false factual belief, one is similarly not blameworthy for possessing a false moral belief. Thus, if a person is ignorant of any fact, moral or factual, a person's actions are excused by her ignorance. To demonstrate his view, Rosen provides the following case:

[317] For Harman's view, see Harman, "Does Moral Ignorance Exculpate?" For Rosen's view, see Rosen, "Culpability and Ignorance." All future references to their respective arguments come from these sources unless indicated otherwise.
[318] Rosen, "Culpability and Ignorance," 64.

> **Ancient Slavery**: In the ancient Near East in the Biblical period the legitimacy of chattel slavery was simply taken for granted. No one denied that it was bad to be a slave, just as it is bad to be sick or deformed. The evidence suggests, however, that until quite late in antiquity it never occurred to anyone to object to slavery on grounds of moral or religious principle. So consider an ordinary Hittite lord. He buys and sells human beings, forces labour without compensation, and separates families to suit his purposes. Needless to say, what he does is wrong. The landlord is not entitled to do these things. But of course he thinks he is.[319]

In **Ancient Slavery**, the Hittite lord has a false moral belief that he is entitled to buy and sell humans. Rosen argues that the Hittite lord is not blameworthy for possessing this belief. Because of this false belief, the Hittite lord engages in buying and selling humans. If he is not blameworthy for possessing this false moral belief, the Hittite lord is similarly not blameworthy for the actions that stem exclusively from this false moral belief.

For Rosen, then, just as factual ignorance can be blameless, moral ignorance can also be blameless. Cases in which people fail to "see through a pervasive and well-protected ideology" in their society are instances of blameless moral ignorance.[320] Conversely, moral and factual ignorance are both blameworthy if they are the upshot of epistemic negligence or recklessness. According to Rosen, moral and factual ignorance exculpate in the exact same way. Actions performed on the basis of *blameless* false moral beliefs are equivalently blameless to actions performed on the basis of false factual beliefs. Actions performed on the basis of *blameworthy* false moral beliefs are equivalently blameworthy to actions performed on the basis of blameworthy false factual beliefs.

Harman take the opposite view. She argues that false factual beliefs and false moral beliefs do not exculpate under the same conditions. She presents the following view:

> **Harman's View**: We are morally obligated to believe the moral truths relevant to our actions (and thus not to believe false moral claims relevant to our actions), and we are often blameworthy for failing to meet these moral obligations, even if we have not been guilty of mismanagement of our beliefs, and even if our ignorance is not motivated. Wrong actions that result from false moral beliefs are not thereby blameless.[321]

Though we do not have an obligation to possess correct factual beliefs, we do have a moral obligation to possess correct moral beliefs. Thus, according to Harman, blameless false moral beliefs do not

---

[319] Rosen, 64–65.
[320] Rosen, 67–68.
[321] Harman, "Does Moral Ignorance Exculpate?," 459.

exist. Rather, any false moral belief is necessarily blameworthy. Since all false moral beliefs are necessarily blameworthy, moral ignorance cannot exculpate.

As demonstrated in the **Cyanide** case, one is not blameworthy for a wrong action that stems exclusively from a false factual belief, as long as the false factual belief is epistemically responsible. According to Harman, however, a person is blameworthy for a wrong action that stems exclusively from a false moral belief, even if the false moral belief is epistemically responsible. To demonstrate how an epistemically responsible false moral belief can be blameworthy, consider the following case, which is presented by Harman:

> **Anti-Same-Sex-Marriage Advocate**: Edgar believes that homosexuality is morally wrong and that the government should not endorse these morally wrongful unions by allowing same sex couples to legally marry. He works hard to prevent same sex marriage from being legal in his state; he is a prominent public intellectual whose advocacy has a real effect. (Edgar is a sophisticated contemporary opponent of same-sex marriage who has no false non-moral beliefs about homosexuals and homosexuality.)[322]

Harman then explains that Edgar's work contributes to preventing the legalization of same-sex marriage, which harms Gary and his same-sex partner. In this case, Edgar possesses a false moral belief, and he performs a wrong action on the basis of the false moral belief. Suppose that Edgar's false moral belief is epistemically responsible. After consulting many people and reflecting on his belief, Edgar arrives at the conclusion that homosexuality is morally wrong. If factual and moral ignorance operate in the same way, then Edgar's false moral belief would excuse his wrong action. Indeed, if factual and moral ignorance exculpate in the same way, then Edgar's wrong action would be just as blameless as Anne mistakenly spooning cyanide into Bill's coffee. Yet, it would be completely understandable, and perhaps even expected, for Gary to resent Edgar, while it would be unreasonable for Bill to resent Anne for her honest mistake.

Edgar's false moral beliefs do not seem to have the same exculpatory power as Anne's false factual beliefs. Even though epistemically responsible false factual beliefs are exculpatory, Edgar's epistemically responsible false moral belief does not exculpate. As Harman explains, a wrong action that results from a false moral belief is not blameless. If a wrong action performed on the basis of a false moral belief is not blameless, then the resulting harm is a blameworthy harm. Thus, we arrive at the conclusion that a wrong action based on a false moral belief is a blameworthy harm.

---

[322] Harman, "Ethics Is Hard! What Follows?," 17–18.

Indeed, Rosen's view is not inconsistent with this conclusion. Rather, he thinks that the conclusion only applies to a subset of false moral beliefs, namely *blameworthy* false moral beliefs. (Harman just thinks that all false moral beliefs are necessarily blameworthy.) For the purpose of clarity and concision in this argument, I exclude cases that Rosen would consider instances of blameless false moral beliefs. Instead, I adopt the agreed upon premise that a wrong action based on a *blameworthy* false moral belief is a blameworthy harm, and we thus arrive at Premise 2.

§5. Incorrect Punishment & Over-Punishment

In this section, I defend the following two premises:

> 3.* A typical case of incorrect punishment is a wrong action based on a blameless false factual belief where the action would have otherwise been permissible had the false factual belief been true.
>
> 4. A typical case of over-punishment is a wrong action based on a blameworthy false moral belief.

If these premises are correct, then, together with Premise 1* and Premise 2, we can reach our initial sub-conclusions.

5.1 Incorrect Punishment

Let's start with Premise 3*, which states that a typical case of incorrect punishment is a wrong action based on a blameless false factual belief where the action would have otherwise been permissible had the false factual belief been true.

In typical cases of incorrect punishment, the state performs a wrong action, namely punishing an innocent person. If punishment is only permissible when a person has committed a crime, then the punishment of an innocent person is impermissible.

The wrong action of punishing an innocent person is often based exclusively on a false factual belief. Let us return to Mazza's case. Mazza's punishment was based on the false factual belief that Mazza robbed and killed the victim. If we assume that the system of punishment is just and that he received the appropriate sentence for a person who commits a robbery-murder, then the only false belief involved is a factual one. As such, the wrong action of incorrect punishment was based on a false factual belief that Mazza had committed the crime. Similarly, in other typical cases of incorrect punishment, the false belief that gives rise to wrong action is often a factual belief that the person committed the crime, when in fact she did not.

Along this dimension, incorrect punishment is similar to the **Cyanide** case discussed above. In both situations, the agents possess all the correct moral beliefs. Anne correctly believes that it is permissible to put sugar in Bill's coffee, and the state correctly believes that it is permissible to punish a person who has committed a robbery-murder. Similarly, in both situations, the agents possess a false factual belief that gives rise to the wrong action. Anne incorrectly believes that she is putting sugar in Bill's coffee, when she is actually putting cyanide in his coffee. As a result, Anne poisons Bill. The state incorrectly believes that Mazza committed the crime, when in fact Mazza is innocent. As a result, the state punishes Mazza. Thus, Mazza's case of incorrect punishment is a wrong action based on a false factual belief.

In both Mazza's case and **Cyanide**, the action would have been permissible if the false factual belief had been true. In the case of incorrect punishment, if the false factual belief had been true—if Mazza had in fact committed the crime—then the state would have been justified in punishing Mazza, and the action would have been permissible. Similarly, if Anne's false factual belief had been true—if the substance in front of her had in fact been sugar—then Anne's action would have been permissible.

One may object, however, that there are instances of incorrect punishment in which there are also false moral beliefs involved. For instance, the state may have false moral beliefs about the types of punishment which are permissible to impose. Suppose that it is never permissible to torture someone as punishment. Suppose further that the state possesses the false moral belief that torture is a permissible punishment. The state then tortures an innocent person both because of a false factual belief about the person's guilt and because of a false moral belief about the permissibility of torture.

I respond that such a case is similar to the **Cyanide & Ricin** case. If the action was going to be impermissible anyways, the false factual belief does not exculpate. In the case of **Cyanide & Ricin**, Anne would have poisoned Bill either way. Since Anne possesses the false moral belief that it is permissible to poison Bill, Anne's resulting action is impermissible. Similarly, if torture is never permitted, the state's action was going to be impermissible, whether the state was torturing a person who had committed a crime or a person who had not committed a crime. Since the state possesses the false moral belief that torture is permissible, the resulting action is not excused by the false factual belief about the criminal history of the person the state tortures.

In a typical case of incorrect punishment, however, the state does not possess false moral beliefs. Instead, the state subjects a person to a punishment that would have been justified had the person committed a crime. As long as the punishment would have been justified had the person been guilty, then the action would have been permissible if the state's belief that the person committed the

crime had been true. Thus, in a typical case of incorrect punishment, the state's moral beliefs about the relevant punishment are otherwise correct.

Alternatively, one may object that there are instances of incorrect punishment in which there are suspicious false factual beliefs involved. For instance, suppose that the state is unsure whether a person committed a crime. Suppose further that, instead of continuing to investigate, the state incorrectly concludes that the person committed the crime. In such a scenario, the state's belief is a false factual belief. Yet, it seems morally dubious to punish on that basis.

I respond that such a case is similar to the **Cyanide & Sugar** case. In that case, Anne knew that there was sugar and cyanide in the kitchen, and that the two containers could be easily confused. Instead of double checking, she spoons the substance in front of her into Bill's coffee. Her false factual belief in that case is epistemically reckless and irresponsible. As a result, the false factual belief is not blameless, and thus does not exculpate. Similarly, if the state is unsure about whether a person committed a crime, it epistemically reckless and irresponsible to believe that the person committed the crime without further investigation into the subject. Thus, the false factual belief is not blameless, and thus does not exculpate.

In a typical case of incorrect punishment, however, the state's belief that a person committed a crime is not reckless. Rather, the state typically reasonably relies on investigators, prosecutors, judges, and juries to determine if a person is guilty. Thus, in a typical case of incorrect punishment, the state's belief about the person's guilt is not epistemically reckless.

To review, a typical case of incorrect punishment is a wrong action based on a blameless false factual belief. It is a wrong action to punish an innocent person, and the action is based on the false factual belief that the person committed a crime. If the person had in fact committed the crime, then the action would have been permissible. Furthermore, in a typical case of incorrect punishment, the state does not have false moral beliefs, since we can assume that it would have been permissible to punish the person had she in fact committed the crime. Moreover, in a typical case of incorrect punishment, the state's false factual belief is epistemically responsible since the state reasonably, though falsely, believes that the person committed the crime. Now that we have defended Premise 3*, we turn to Premise 4.

5.2 Over-Punishment

Typical cases of over-punishment are distinct from typical cases of incorrect punishment in that there may be no false factual beliefs involved. Premise 4 states that over-punishment is a wrong action based

on a blameworthy false moral belief. In cases of over-punishment, the state performs a wrong action, namely harming a person beyond the limits of justified punishment. If harming a person who has committed a crime is only permissible within the limits of justified punishment, then harming a person beyond those limits is a wrong action.

The wrong action involved in over-punishment, however, is based on a false moral belief. In Andrade's case, for instance, the state knows all of the relevant factual facts. The state correctly believes that Andrade had committed the crime of shoplifting $153 worth of children's videos. The state also correctly believes that Andrade had committed two other crimes. These are correct factual beliefs. The relevant false moral belief, however, is that it is permissible to sentence a person to at least 50 years of incarceration for committing three non-violent minor thefts. Thus, the wrong action is a result of a false moral belief, since all of the factual facts of the case are apparent and correct. Again, I remain agnostic regarding what justifies punishment and its limits. The false moral belief in any case of over-punishment is simply the belief that it is permissible to impose a given treatment that is actually impermissible to impose since it is beyond the limits of justified punishment.

Since a typical case of over-punishment involves a false moral belief, we must determine whether the relevant false moral belief can render the resulting wrong action a blameless mistake. Indeed, one could argue that it is hard to know the proper limits of punishment. As such, perhaps the state is not blameworthy for possessing a false moral belief about the limits of permissible punishment.

I have four responses. First, if we adopt Harman's view, then the state's false moral belief is necessarily blameworthy since there are no blameless false moral beliefs. Second, even if some false moral beliefs could be blameless, false moral beliefs do not necessarily exculpate in the same way that false factual beliefs can. As we saw in the **Anti-Same-Sex-Marriage Advocate** case, Gary's resentment towards Edgar would not be misplaced, and Edgar is not excused by his false moral belief. If anything, Edgar's false more belief may render him even more blameworthy. Similarly, it would be reasonable and expected for Andrade to resent the state for its wrong action. Again, Andrade's potential resentment towards the state is importantly disanalogous from Bill's resentment towards Anne. Anne is excused by her false belief, while the state is not. The state's belief about the proper limits of punishment, therefore, is a non-exculpatory false moral belief. Since the false moral belief is non-exculpatory, the state's resulting wrong action is a blameworthy harm.

Third, even according to Rosen, it seems that the state's false moral belief in Andrade's case would not qualify as blameless moral ignorance. Rosen's central case, **Ancient Slavery**, is an example of an agent acting within a context in which everyone in the society possesses the same false moral

belief. He states that it would have taken a "moral genius" to recognize that slavery is wrong in that context.[323] Whether or not his case is historically accurate, it seems that our example of a false moral belief is already fairly disanalogous. When Andrade was sentenced in California in 1995, it would not have taken a moral genius to recognize that a 50-year sentence is likely an unjustified punishment for three instances of non-violent petty theft.

Moreover, the very existence of different theories of punishment indicates that there is not a "pervasive and well-protected" cultural ideology.[324] Indeed, Rosen contrasts the **Ancient Slavery** case with American race-based slavery in the 1800s.[325] In that context, one would not be able to appeal to blameless moral ignorance to justify slavery. It was widely known that many people opposed the practice and correctly believed that slavery is a grave moral wrong. Similarly, beliefs about the limits of punishment cannot be construed as blameless moral ignorance when there is widespread disagreement about and opposition to certain practices. Indeed, many people criticized California's three strikes law and its application in cases like Andrade's when it was passed in 1995.[326]

Fourth, if the state is unsure about the limits of punishment, then it would be a case of moral uncertainty, not blameless moral ignorance. Alex Guerrero, among others, argues that moral uncertainty is distinct from moral ignorance.[327] Moral ignorance may arise either when an agent is completely unaware of a certain moral fact or when an agent believes the opposite of the moral fact. Moral uncertainty occurs when the agent has thought about the issue but is unsure what to believe. Thus, if the state is unsure about the proper limits of punishment, the state is morally uncertain—not morally ignorant. Moral uncertainty, however, cannot exculpate a wrong action in the same way. Rather, as Guerrero argues, in the face of uncertainty, an agent should act with caution.

Though it may be the case that some punishment-related moral facts are difficult to discern, there will be plenty of punishment-related moral facts that are quite clear. For instance, it may be difficult to figure out whether a person deserves 10 months or 10.5 months of incarceration for a specific crime. Indeed, this epistemic concern has been raised across the board for many theories of punishment.[328] Yet, in many cases, it is clear that a punitive practice is obviously problematic. For

---

323 Rosen, "Culpability and Ignorance," 66.
324 Rosen, 67–68.
325 Rosen, 65. Rosen uses the contrast to demonstrate that ancient slavery was "simply taken for granted."
326 For criticism of the law from 1995 when it was first passed, see, e.g., Press, "25 Years for a Slice of Pizza." For criticism in the following years, see, e.g., ACLU, "10 Reasons to Oppose '3 Strikes, You're Out.'"
327 Guerrero, "Don't Know, Don't Kill: Moral Ignorance, Culpability, and Caution."
328 See, e.g., Kolber, "The Comparative Nature of Punishment"; Husak, "Already Punished Enough."

instance, it seems clearly wrong for the state to torture a person who shoplifts a candy bar.[329] Thus, many of the state's beliefs about the limits of permissible punishment cannot be protected by an appeal to blameless moral ignorance. In Andrade's case and other cases of over-punishment, the wrong action is based on a blameworthy false moral belief.

We have established that wrong actions based on a false factual belief may be blameless mistakes. A typical case of incorrect punishment is a wrong action based on a false factual belief that the person did in fact commit the crime. Thus, a typical case of incorrect punishment is a blameless mistake (SC1). In contrast, a blameworthy false moral belief cannot exculpate. A typical case of over-punishment is a wrong action based on a blameworthy false moral belief that a given practice or treatment is permissible when it is actually impermissible. Since that belief is non-exculpatory, the resulting harm is blameworthy. Thus, a typical case of over-punishment is a blameworthy harm (SC2).

Let's review the established premises in the **Compensation for Over-Punishment** argument:

1.* A wrong action based on a blameless false factual belief is a blameless mistake if and only if the action would have been permissible had the false factual belief been true.

2. A wrong action based on a blameworthy false moral belief is a blameworthy harm.

3.* A typical case of incorrect punishment is a wrong action based on a false factual belief where the action would have otherwise been permissible had the false factual belief been true.

4. A typical case of over-punishment is a wrong action based on a blameworthy false moral belief.

With these four premises, we arrive at the first two sub-conclusions.

SC 1: A typical case of incorrect punishment is a blameless mistake. (1, 3)

SC 2: A typical case of over-punishment is a blameworthy harm. (2, 4)

Now, let's turn to our final premise about compensation.

§6. Compensation

In this section, I defend the last premise:

5. The argument for compensation for a blameworthy harm is stronger than the argument for compensation for a blameless mistake.

To see how compensation tracks the wrongfulness of a harm, let us again return to the **Cyanide** case. In **Cyanide**, Anne's belief that she is spooning sugar into Bill's coffee is a false factual belief. Anne

[329] Indeed, all of the standard theories of punishment would agree that such a punishment would be impermissible.

possesses all the relevant true moral beliefs, namely that it is impermissible to poison Bill. In contrast, consider another version of the case with false moral beliefs:

> **Intentional Cyanide**: Anne spoons cyanide into Bill's coffee, and she knows it is cyanide: she believes that it is permissible to poison Bill.

In **Intentional Cyanide**, Anne possesses all the relevant true factual beliefs. She is aware that she is spooning cyanide into Bill's coffee. Yet, Anne does not possess all the relevant true moral beliefs because she incorrectly believes that it is permissible to poison Bill.

In both cases, Bill experiences the same harm, namely that he is poisoned. In **Cyanide**, however, Anne's action is based on a false factual belief. Since the false factual belief is exculpatory, Anne's action is excused, rendering her wrong action a blameless mistake. In **Intentional Cyanide**, Anne's action is not excused. Since Anne's action is based on a false moral belief, the false moral belief is non-exculpatory, so her wrong action remains a blameworthy harm.

Would Anne owe compensation to Bill in **Cyanide** and **Intentional Cyanide**? In **Intentional Cyanide**, Anne clearly owes Bill compensation since Anne wrongfully harms Bill. In **Cyanide**, however, it is less clear. On the one hand, compensation could be owed only when a person's action is a blameworthy harm.[330] Since Anne's action in **Cyanide** is a blameless mistake, compensation might not be owed. On the other hand, compensation could be owed for any harm, since the harmed person is worse off.[331] According to this view, the compensation is based on Bill's loss, not the culpability of Anne's action. The goal of the compensation, then, is to restore Bill to his original position. Thus, for a proponent of this view it would not matter whether Anne pays or someone else pays.[332] Indeed, a proponent of this view may support the "beneficiary pays" model.[333] If a person benefits from the harm, then, as the beneficiary, she should compensate the person who is worse off in order to attempt to restore everyone to their previous standing. Thus, it is far less clear if Anne would owe compensation to Bill for her blameless mistake arising from false factual belief.

---

[330] A version of this view is presented by Joel Feinberg. He explains the difference in compensatory duties in the context of rights infringements (blameless harm) and rights violations (blameworthy harm). Regarding a blameless harm he states, "If the other had no right that was infringed in the first place, one could hardly have a duty to compensate him. Perhaps he would be an appropriate object of your sympathy or patronage or charity, but those are quite different from compensation" (Feinberg, "Voluntary Euthanasia and the Inalienable Right to Life," 102).

[331] A version of this view is defended by Judith Jarvis Thompson. In response to Feinberg, she argues that compensation is owed when a person is harmed, even if a right has not been infringed (i.e., for both blameworthy and blameless harms). See Thomson, "Rights and Compensation."

[332] Following Thomson, Davis clarifies that Thomson's view includes the possibility that the compensation is not owed by the person who causes the harm. See Davis, "Rights and Moral Theory."

[333] For a defense of the "beneficiary pays" model, see, e.g., Butt, "On Benefiting from Injustice."

If compensation is only owed when the harm is blameworthy, then Anne would not owe compensation in **Cyanide**, but she would owe compensation in **Intentional Cyanide**. If compensation is owed in all cases of harm, then compensation would be owed in both **Cyanide** and **Intentional Cyanide**. There is no plausible view according to which Anne owes compensation in **Cyanide** and does not owe compensation in **Intentional Cyanide**.

Anne's obligation to compensate Bill is stronger in the **Intentional Cyanide** case. If Anne had limited resources, it is clear that Anne would need to first provide compensation for her blameworthy harm in **Intentional Cyanide**, and only after would she be able to provide compensation for her blameless mistake in **Cyanide**. Indeed, it is quite plausible that Anne may not owe any compensation for her blameless mistake in **Cyanide**. Thus, we see that there is a stronger argument for compensation in cases of blameworthy harms than in cases of blameless mistakes.

Thus, we have established that the argument for compensation for a blameworthy harm is stronger than the argument for compensation for a blameless mistake (Premise 5). We have also established that a typical case of incorrect punishment is an instance of a blameless mistake (SC1), and that a typical case of over-punishment is an instance of a blameworthy harm (SC2). Given these three premises, we arrive at our surprising conclusion: there is a stronger argument for compensation for a typical case of over-punishment than for compensation for a typical case of incorrect punishment.

One may object, however, that perhaps the argument for compensation for over-punishment is weaker than the argument for compensation for incorrect punishment, since a person who commits a crime could have known about the consequences for committing that crime. Since the over-punished person could have expected the over-punishment when breaking the law, perhaps compensation is less clearly owed.

I respond, however, that even if people know the consequences of their actions, a blameworthy harm is still more problematic than a blameless mistake. Consider the following case:

> **Cyanide Threat**: Anne wants Bill to loan her $500. She tells him that if he does not loan her the money, she will poison his coffee. Bill does not loan Anne the money, so Anne puts cyanide in Bill's coffee.

In **Cyanide Threat**, Bill has all of the information about what will happen if he does not loan Anne the $500 that she has requested. Yet, the fact that Anne first threatens Bill does not appear to make Anne's wrong action less blameworthy. If anything, it does the opposite. If we compare **Cyanide Threat** with our original **Cyanide** case, we again see that there is a stronger argument for compensation in **Cyanide Threat** than in **Cyanide.**

Thus, we again arrive at the conclusion that there is a stronger argument for compensation in cases of over-punishment than for compensation in cases of incorrect punishment. In the coming section, we explore non-standard examples of over-punishment and incorrect punishment.

§7. Typology of Harmful Punishments

Thus far, we have discussed a typical case of incorrect punishment and a typical case of over-punishment. What about atypical cases of incorrect punishment and atypical cases of over-punishment?

I argue that cases with only blameless false factual beliefs are blameless mistakes, while cases with at least one blameworthy false moral belief are blameworthy harms. To demonstrate this, consider the following chart of the possible scenarios and the discussion below.

	True Moral Beliefs	False Moral Beliefs
True Factual Beliefs	1: No Compensation (No Wrong)	3: Over-Punishment (Blameworthy Harm)
False Factual Beliefs	2: Incorrect Punishment (Blameless Mistake)	4: Incorrect Over-Punishment (Blameworthy Harm)

Type 1 represents standard cases of punishment. Suppose that a person is guilty and that the state possesses all of the relevant true factual and moral beliefs. If the state then harms the person in accordance with the state's true moral beliefs, the state does not wrong the person. Thus, no compensation is owed. Type 1 also represents standard cases of non-punishment. Suppose that a person is innocent and that the state possesses all the relevant true factual and moral beliefs. In such a case, the state would not harm the innocent person in accordance with the state's true moral belief that it is impermissible to punish innocent people. Thus, there is no harm.

Type 2 represents cases of incorrect punishment. As discussed above, typical cases of incorrect punishment involve false factual beliefs about the person's guilt and correct moral beliefs about permissible punishment. In such cases, the false factual belief exculpates, and thus the harm is a blameless mistake.

Type 2 also includes atypical cases such as those in which a person had in fact committed a crime, but there are false factual beliefs involved. Consider the following case:

Different Crime: Celia is guilty of manslaughter. The state has the false factual belief that Celia is guilty of murder. The state has the correct moral belief that a 10-year sentence is the longest permissible punishment for murder. The state also has the correct moral belief that a three-year sentence is the longest permissible punishment for manslaughter. The state gives Celia a 10-year sentence.

Though the state correctly believes that Celia committed a crime, the state incorrectly believes that Celia committed a murder. The state possesses all of the relevant true moral beliefs, since the state's punishment scheme is justified. In this case, the false factual belief about what exactly Celia did (or intended) does, in fact, exculpate. The state's action would have been permissible had the false factual belief been true. If Celia had in fact committed a murder, the state would have permissibly imposed a 10-year sentence. Thus, this case is analogous to a typical case of incorrect punishment. In an instance of incorrect punishment, the state has a false belief about *who* committed the crime, while in **Different Crime**, the state has a false belief about *which* crime was committed. Either way, the state believes that a person committed a crime that she in fact did not commit. Regardless of whether a person is guilty or innocent, the wrong action in a Type 2 case is a blameless mistake.

Type 3 represents cases of over-punishment. As discussed, typical cases of over-punishment involve true factual beliefs about the person convicted and the crime committed. The state has false moral beliefs about permissible punishment. Thus, the state wrongfully harms the individual. Given that in Type 3 the state possesses true factual beliefs, it is hard to construe a case in which the state knowingly punishes an innocent person. Yet, suppose that the state possesses the true factual belief that a person is innocent, as well as the false moral belief that it is permissible to punish innocent people. In such a case, the state's action would also be a blameworthy harm.

Type 4 represents cases in which the state has both false moral beliefs and false factual beliefs. We previously considered an instance of Type 4 when we discussed punishing an innocent person with torture. Consider the following case:

Torturing the Innocent: Celia is innocent. The state possesses the false factual belief that Celia committed a murder. The state also has the false moral belief that torture is permissible when a person commits a murder. The state tortures Celia.

If it is never permissible to torture someone as punishment, then state possesses the false moral belief that torture is a permissible punishment for certain crimes. Since Celia is innocent, the state also has a false factual belief that Celia committed a crime. The state then harms Celia by torturing her. The state acts in accordance with both a false factual belief about Celia's guilt and a false moral belief about

the permissibility of torture. This case is similar to the **Cyanide & Ricin** case, since Anne spoons cyanide into Bill's coffee both because she incorrectly thinks it is permissible to poison Bill and because she incorrectly thinks she is spooning ricin. If the action would have been wrongful anyways, then the false factual belief does not exculpate. If the torture would have been wrongful even if Celia had been guilty, then the resulting action is not excused by the false factual belief about Celia's criminal history. As a result, the state's action is a blameworthy harm.

Type 4 can also occur when a person is guilty. Suppose the state correctly believes that the person is guilty but also possesses another relevant false factual belief and a false moral belief. Consider a modified version of **Torturing the Innocent**:

> **Torturing the Guilty**: Celia is guilty of manslaughter. The state has the false factual belief that Celia is guilty of murder. The state also has the false moral belief that torture is a permissible punishment for murder. The state tortures Celia.

If it is never permissible to torture someone as punishment, then state has the false moral belief that torture is a permissible punishment for certain crimes. The state correctly believes that Celia is guilty. The state incorrectly believes that Celia committed a murder. Thus, the state possesses a relevant false factual belief. The state then harms Celia by torturing her. As in **Torturing the Innocent**, if the action would have been wrongful anyways, the false factual belief does not exculpate. Since torturing anyone is wrongful, the false factual belief about which crime Celia committed does not turn the torture into a blameless mistake. Thus, cases of Type 4—regardless of whether the person is innocent or guilty—are instances of a blameworthy harm.

There are atypical cases of over-punishment and atypical cases of incorrect punishment with different types of false beliefs. When the false belief is only a factual one, it is a blameless mistake like a typical case of incorrect punishment. When at least one of the relevant false beliefs is a moral one, then it is a blameworthy harm like a typical case of over-punishment.

§8. The State's Right to Err

Thus far, we have drawn on the literature regarding the exculpatory power of false beliefs to identify and evaluate the false beliefs, wrong actions, and compensatory duties involved in cases of incorrect punishment and over-punishment. One may be concerned, however, that there is an important distinction between cases in which the agent is an individual and cases in which the agent is the state. There may be a variety of ways in which the state functions differently as an agent than do individuals, both descriptively and normatively.

In this section, I evaluate the state's false beliefs, wrong actions, and compensatory duties in the context of punishment and state legitimacy. I argue that the **asymmetry conclusion** primarily still applies, namely that there is a stronger argument for compensation for over-punishing the guilty than for compensation for incorrectly punishing the innocent. On certain views and under certain conditions, the **symmetry conclusion** may be more fitting, namely that the same argument applies to compensation both for over-punishment and for incorrect punishment. Thus, compensation is due for incorrect punishment if and only if compensation is due for over-punishment. Regardless of whether one adopts the asymmetry conclusion or the symmetry conclusion, compensation is never only owed in cases of incorrect punishment. Rather, if compensation is owed in cases of incorrect punishment, compensation is certainly owed in cases of over-punishment.

8.1 Beliefs and Actions

The literature on the exculpatory power of false beliefs focuses on individuals. The state, however, is not an individual but rather a collection of individuals. This raises the questions: Can groups possess beliefs? If so, can groups act on the basis of those beliefs?

Let us start with the question of group belief. In the literature on collective epistemology, the debate centers around how collective beliefs are formed and when collective beliefs are justified. The starting assumption, therefore, is that a group can have a belief. Indeed, governments are often the classic example of a group that possesses a belief that leads to an action. Jennifer Lackey, for instance, in the introduction to *Essays on Collective Epistemology* uses the following example of a group belief:

> If the Bush Administration justifiably believed that Iraq did not have weapons of mass destruction, then not only did the Administration lie to the public in saying that it did, but it is also fully culpable for the hundreds of thousands of lives needlessly lost in the Iraq war.[334]

In this example, Lackey assumes (a) that a group of government officials can possess a belief, (b) that the group of government officials can act on the basis of a belief, and (c) that the group of government officials can be culpable for its actions.

One may be concerned that moral beliefs and factual belief function differently in the context of collective belief. Perhaps, a group can possess a factual belief, as demonstrated in Lackey's Bush administration example, but a group cannot possess a moral belief.

[334] Lackey, "Introduction," 1.

I respond that there is no obvious reason to differentiate between moral beliefs and factual beliefs in this context. If a group can possess a collective factual belief, there is no reason to assume that the group cannot possess a collective belief about what the group *should* do. If the group can possess a belief about what the group should do, then it appears that the group can possess moral beliefs, as well.

Indeed, groups must hold some moral beliefs about what the group should do in order to act. Suppose that, as Lackey's example suggests, the Bush Administration believed that Iraq did not have weapons of mass destruction, then some further belief would be needed to explain the Iraq war. For instance, the Bush Administration may have possessed the following belief: even if Iraq did not have weapons of mass destruction, we should still say that there are weapons of mass destruction in order to gain popular support for the invasion. With that (false) moral belief, the resulting action—namely invading Iraq—can be explained.

Not only can groups possess moral beliefs, but they can also be held responsible for moral actions. Christian List and Philip Pettit argue that groups can be held responsible under the following conditions:

1) The group agent faces a normatively significant choice, involving the possibility of doing something good or bad, right or wrong.
2) The group agent has the understanding and access to evidence required for making normative judgments about the options.
3) The group agent has the control required for choosing between the options.[335]

Notably, in List and Pettit's conditions, the group agent must have access to evidence required for making normative judgments. Since moral beliefs are necessary for normative judgments, the group must be able to possess moral beliefs. With those moral beliefs and the resulting actions, a group may be held responsible if the group meets List and Pettit's conditions.

On the basis of List and Pettit's conditions, it appears that a group of state officials who administer a punishment system can be held responsible for its actions. The group faces normatively significant choices, namely whether they should impose punishments in different cases. The group has access to evidence, including police reports, prosecutorial choices, and jury verdicts. The group may also possess the moral belief that they should impose punishment when and only when there is a guilty verdict. Finally, the group of state officials can control whether or not punishment is imposed.

[335] List and Pettit, *Group Agency: The Possibility, Design, and Status of Corporate Agents*, 158.

If the relevant agent is a group of state officials who administer the punishment system, then it seems directly analogous to individuals. The group can believe, act, and become culpable, just as individuals can believe, act, and become culpable. Thus, it seems that the asymmetry conclusion stands even though the agent is a group of state officials instead of an individual.

Notably, a jury plays different roles in instances of incorrect punishment than in instances of over-punishment. Like government groups, juries are often cited as a group that can believe and act.[336] In the context of over-punishment, the jury transmits a true belief, namely that a person committed a certain crime. In the context of incorrect punishment, however, the jury transmits a false belief, namely that a person committed a crime when she did not.

In cases of incorrect punishment, the jury's belief is the origin of the state's false belief. The transmission of the belief from the jury to the state can be thought of as testimony. If the jury can be considered a reliable source, the state's false belief is a false factual belief adopted through testimony from a reliable source. Thus, it eliminates the question of whether the state acted epistemically recklessly or irresponsibly when forming the false factual belief, since standardly factual testimony from a reliable source is epistemically reasonable to adopt. Thus, since the jury is the group that generates and transmits the false belief, the false factual testimony further emphasizes the potential blamelessness of the state's adopted false factual belief.

When it comes to false moral beliefs, however, the jury is not involved. The jury only provides factual testimony. Since the jury's factual testimony is correct in cases of over-punishment, the jury excuse is not available.[337] The discussion of the role of juries further emphasizes why false beliefs in instances of incorrect punishment may be blameless, while false beliefs in instances of over-punishment may be blameworthy. Thus, the asymmetry between incorrect punishment and over-punishment remains.

One may be concerned, however, that the state is a special type of group with different licenses and constraints. If the state is legitimate, then it has certain powers to coerce people. Thus, certain actions, which would otherwise be impermissible for an individual to perform, may be permissible for the state to perform. I address this concern in the coming section.

---

[336] Though the assumption is present throughout the legal epistemology literature, Michael Pardo argues directly for the claim that juries possess the power to believe, know, and act. See Pardo, "Group Agency and Legal Proof; or, Why the Jury Is an 'It.'"

[337] An exception to this may be cases that are eligible for the death penalty in the US. Capital cases are the only type of criminal case in which the jury gets to choose the penalty (life without parole or death) in a sentencing phase after conviction. See, e.g., Haney, Sontag, and Costanzo, "Deciding to Take a Life."

8.2 Legitimacy

In cases of both incorrect punishment and over-punishment, people are harmed by the state. When a state exercises its power, it inevitably makes mistakes. Any plausible theory of legitimacy, therefore, has to account for the state mistakenly applying good policy. If a theory of legitimacy requires that the state never makes mistakes, then likely every state would be illegitimate. Rather, due to the state's special position, it may be the case that the state's action is permissible, even when it harms a citizen. Indeed, political theorists have recognized this phenomenon as part of a state's legitimacy.[338]

In a legitimate state, when a person is over-charged on her taxes, the person is harmed since she is forced to pay more than she owes. Yet, the state's action is permissible since the state possesses the leeway to make mistakes. This does not, however, give the state free license to harm people. For instance, the state cannot permissibly commit genocide against its own people.

Incorrect punishment is structurally analogous to a case in which a person is over-charged on her taxes. Both are instances in which good policy, with proper safeguards, is implemented incorrectly. In the context of over-charging taxes, the tax code is a good policy, and generally the tax collectors are careful to ensure that people are not over-charged. When a person is over-charged, the state is mistakenly applying a good policy. In the context of incorrect punishment, the policy (punish people who commit crime x with punishment y) is also good policy. There are also safeguards in place, such as jury trials, to reduce errors. Yet, the state makes a mistake about whether the person committed crime x, and thus erroneously applies a good policy.

The question, then, is whether the state permissibly harms people in cases of over-punishment. Whether over-punishment is a case of permissible harming or impermissible wronging will depend on a number of factors. Political theorists have suggested that different factors that impact the permissibility of the state's wrong action, including the process for creating a policy, the degree of harm inflicted, and the distribution of harm.[339] A thorough analysis of how different types of over-punishment would map onto to different theories of legitimacy is beyond the scope of this paper. But, if the over-punishment is indeed a wrong, then there is still an asymmetry, since incorrect punishment

---

[338] See, e.g., Wellman, "The Space between Justice and Legitimacy"; Viehoff, "Legitimacy as a Right to Err."

[339] For instance, according to one view, if there is a tremendous amount of benefit from the over-punishment, then the over-punishment would be an instance of permissible harming. If, however, the harm was disproportionally large relative to the benefit it provides, then the over-punishment would be an instance of impermissible wronging. See, e.g., Wellman, "The Space between Justice and Legitimacy." According to another view, if the over-punishment policy is created as a result of thin disagreement in a democratic process, then the resulting harm is an instance of permissible harming. See, e.g., Valentini, "Justice, Disagreement and Democracy." Thus, whether an instance of over-punishment is an impermissible wrong or a permissible harm will depend on the details of the situation.

is a permissible harm. If the over-punishment is also a permissible harm, then it is in the same category as incorrect punishment. This would lead us to the **symmetry conclusion**: compensation for one is due if and only if compensation for the other is due.

8.3 Compensatory Duties

Thus far, we have seen that cases of over-punishment may be permissible harms or impermissible wrongs, while cases of incorrect punishment are cases of permissible harms. What, if any, compensation does the state owe for permissible harms and impermissible wrongs?

Political theorists, including Daniel Viehoff, have argued that part of the state's right to err is that the state does not possess compensatory duties for harming people under certain conditions.[340] Though the state may harm people, the right to err does not cover all harms that the state imposes. Viehoff, the primary champion of the state's right to err, presents the following limit to the right: "The errors that the right to err covers are, even though morally regrettable, not ones for which we can morally criticize or blame the person making them."[341] Viehoff goes on to argue that the right to err importantly dispenses the state from compensatory duties in the relevant cases.

If we apply Viehoff's view of the right to err to our context, we see that the state would not need to provide compensation in cases of incorrect punishment. The right to err specifically covers "morally regrettable" mistakes, such as convicting the wrong person. Yet, the state would likely need to provide compensation for cases of over-punishment. The distinction that Viehoff draws follows the difference between **Cyanide** and **Intentional Cyanide**. In **Cyanide**, Anne's action is a morally regrettable mistake. Though Bill is harmed, Anne would not be the rightful target of resentment and blame. In **Intentional Cyanide**, however, it would certainly be expected that Bill would resent and blame Anne for purposefully poisoning him. Similarly, Edgar's advocacy against same-sex marriage would also be worthy of moral criticism and blame. In **Intentional Cyanide** and **Anti-Same-Sex-Marriage Advocate**, the actions are not "morally regrettable" mistakes but rather blameworthy harms. Similarly, Andrade and others subjected to over-punishment under the three strikes law would likely resent the state for the harm imposed. Many criticized California and other jurisdictions for sentencing people to life without parole on a third strike.[342] Though not all cases of over-punishment will necessarily fall into this category, any instance of over-punishment that is a harm worthy of moral

[340] For a defense of the state's right to err, see Viehoff, "Legitimacy as a Right to Err."
[341] Viehoff, 186.
[342] See, e.g., Press, "25 Years for a Slice of Pizza."

criticism and blame would fall into this category. If so, then the over-punishment falls beyond the protection of the state's right to err. Thus, in typical cases of over-punishment, unlike typical cases of incorrect punishment, the state can be criticized for unjust policies. The right to err thus further reinforces the asymmetry conclusion.

One could object that the state should not use its resources for compensating people for its mistakes and wrongs. Instead, the state should allocate its resources towards alleviating hunger or developing life-saving medical treatments, for instance. Such an objection would likely not apply on an individual level. If Cara wrongfully harms Daniel and owes Daniel compensation, Cara cannot simply say that the compensation would be best used in another way and that she is therefore not required to compensate Daniel.

On the state level, however, it is possible that the state has to balance different duties to different citizens. The state may have a duty to alleviate their hunger and compensate people who the state has harmed. Indeed, it is likely the case that the duty to alleviate hunger always comes before compensating people who the state has harmed.

Even if the state has other duties that come before compensating individuals who the state has harmed, there is still a stronger argument for compensation for over-punishment. If the state is planning on using any resources to compensate people who have been harmed, it should start by compensating those who have been wrongfully harmed. Only afterwards, if there are still resources available, should the state compensate those who are harmed by blameless mistakes.

Conclusion

In this paper, we started with five fairly uncontroversial assumptions. Drawing primarily on the literature regarding the exculpatory power of false beliefs, we found that compensation for incorrect punishment is actually less defensible than compensation for over-punishment. As such, if one supports compensating innocent people who are incorrectly punished, then one should certainly support compensation for those who are over-punished.

Bibliography

28 U.S. Code § 1495 - Damages for unjust conviction and imprisonment; claim against United States (1948).

28 U.S. Code § 2513 - Unjust Conviction and Imprisonment (1948).

ACLU, ed. "10 Reasons to Oppose '3 Strikes, You're Out.'" American Civil Liberties Union, March 17, 2002. http://www.aclu.org/documents/10-reasons-oppose-3-strikes-youre-out.

Armbrust, Shawn. "When Money Isn't Enough: The Case for Holistic Compensation of the Wrongfully Convicted Note." *American Criminal Law Review* 41, no. 1 (2004): 157–82.

Bedau, Hugo Adam. "The Minimal Invasion Argument Against the Death Penalty." *Criminal Justice Ethics* 21, no. 2 (June 1, 2002): 3–8. https://doi.org/10.1080/0731129X.2002.9992122.

Brooks, Thom. "Retributivist Arguments against Capital Punishment." *Journal of Social Philosophy* 35, no. 2 (2004): 188–97. https://doi.org/10.1111/j.1467-9833.2004.00224.x.

Butt, Daniel. "On Benefiting from Injustice." *Canadian Journal of Philosophy* 37, no. 1 (2007): 129–52.

Chemerinsky, Erwin. "Cruel and Unusual: The Story of Leandro Andrade." *Drake Law Review* 52, no. 1 (2003): 1–24.

Chunias, Jennifer L., and Yael D. Aufgang. "Beyond Monetary Compensation: The Need for Comprehensive Services for the Wrongfully Convicted Essays from the Innocence Project." *Boston College Third World Law Journal* 28, no. 1 (2008): 105–28.

Clow, Kimberley A., Isabella M. Blandisi, Rose Ricciardelli, and Regina A. Schuller. "Public Perception of Wrongful Conviction: Support for Compensation and Apologies." *Albany Law Review* 75, no. 3 (2011): 1415–38.

Davis, Nancy. "Rights and Moral Theory: A Critical Review of Judith Thomson's Rights, Restitution, and Risk." Edited by Judith Jarvis Thomson and William Parent. *Ethics* 98, no. 4 (1988): 806–26.

Faridi, Muhammad U., Hillel Hoffman, and Paul A. Montuori. "Undoing Time: A Proposal for Compensation for Wrongful Imprisonment of Innocent Individuals." *Western New England Law Review* 34, no. 1 (2012): 1–56.

Feinberg, Joel. "Voluntary Euthanasia and the Inalienable Right to Life." *Philosophy & Public Affairs* 7, no. 2 (1978): 93–123.

Guerrero, Alexander. "Don't Know, Don't Kill: Moral Ignorance, Culpability, and Caution." *Philosophical Studies* 136 (2007): 59–97.

Haney, Craig, Lorelei Sontag, and Sally Costanzo. "Deciding to Take a Life: Capital Juries, Sentencing Instructions, and the Jurisprudence of Death." *Journal of Social Issues* 50, no. 2 (1994): 149–76. https://doi.org/10.1111/j.1540-4560.1994.tb02414.x.

Harman, Elizabeth. "Does Moral Ignorance Exculpate?" *Ratio* 24, no. 4 (2011): 443–68.

———. "Ethics Is Hard! What Follows?," July 7, 2014. http://www.princeton.edu/~eharman/Ethics%20Is%20Hard%20020714%20For%20Web.pdf.

Husak, Douglas. "Already Punished Enough." *Philosophical Topics* 18, no. 1 (1990): 79–99.

Jaffe, Ina. "Cases Show Disparity of California's 3 Strikes Law." *NPR*, October 30, 2009, sec. Law. https://www.npr.org/templates/story/story.php?storyId=114301025.

Janse, Alejandra Marquez, and Ari Shapiro. "For the Exonerated, Compensation Is a Battle for Stability and Dignity." *NPR*, January 11, 2023, sec. Law. https://www.npr.org/2023/01/11/1147443227/for-the-exonerated-compensation-is-a-battle-for-stability-and-dignity.

Jasiński, Wojciech, and Karolina Kremens, eds. *Compensation for Wrongful Convictions: A Comparative Perspective*. London: Routledge, 2023. https://doi.org/10.4324/9781003229414.

Kahn, Daniel S. "Presumed Guilty until Proven Innocent: The Burden of Proof in Wrongful Conviction Claims under State Compensation Statutes." *University of Michigan Journal of Law Reform* 44, no. 1 (2010): 123–68.

Kaiser, H. Archibald. "Wrongful Conviction and Imprisonment: Towards an End to the Compensatory Obstacle Course." *Windsor Yearbook of Access to Justice* 9 (1989): 96–153.

Karaffa, Kerry M., Jaimie Page, and Julie M. Koch. "Compensating the Innocent: Perceptions of Exonerees' Deservingness to Receive Financial Compensation for Wrongful Convictions."

Criminal Justice Policy Review 28, no. 7 (2017): 710–32. https://doi.org/10.1177/0887403415607049.

Kolber, Adam J. "The Comparative Nature of Punishment." *Boston University Law Review* 89, no. 5 (2009): 1565–1608.

Lackey, Jennifer. "Introduction." In *Essays in Collective Epistemology*, edited by Jennifer Lackey, 1–8. Oxford University Press, 2014. https://doi.org/10.1093/acprof:oso/9780199665792.003.0001.

List, Christian, and Philip Pettit. *Group Agency: The Possibility, Design, and Status of Corporate Agents*. Oxford University Press, 2011. https://doi.org/10.1093/acprof:oso/9780199591565.003.0002.

Madrigal, Andrew J., and Robert J. Norris. "The Good, the Bad, and the Uncertain: State Harm, the Aftermath of Exoneration, and Compensation for the Wrongly Convicted." *Critical Criminology* 30, no. 4 (December 1, 2022): 895–913. https://doi.org/10.1007/s10612-022-09656-7.

Nathanson, Stephen. *An Eye for an Eye: The Immorality of Punishing by Death*. Rowman & Littlefield, 2001.

Nowotny, Jordan, Amy Shlosberg, and Thomas McAndrew. "Understanding Public Views of Wrongful Conviction Frequency and Government Responsibility for Compensation: Results from a National Sample." *Criminal Justice Policy Review* 34, no. 2 (March 1, 2023): 140–60. https://doi.org/10.1177/08874034221106747.

Otterbourg, Ken. "National Registry of Exonerations: Anthony Mazza." University of Michigan Law School, June 13, 2021. https://www.law.umich.edu/special/exoneration/Pages/casedetail.aspx?caseid=5975.

Pardo, Michael S. "Group Agency and Legal Proof; or, Why the Jury Is an 'It.'" *William & Mary Law Review* 56, no. 5 (2015): 1793–1858.

Press, The Associated. "25 Years for a Slice of Pizza." *The New York Times*, March 5, 1995, sec. U.S. https://www.nytimes.com/1995/03/05/us/25-years-for-a-slice-of-pizza.html.

Quirk, Hannah. "Compensation for Wrongful Convictions in England and Wales." In *Compensation for Wrongful Convictions*, edited by Wojciech Jasiński and Karolina Kremens, 8–24. Routledge, 2023.

Roberts-Cady, Sarah. "Against Retributive Justifications of the Death Penalty." *Journal of Social Philosophy* 41, no. 2 (2010): 185–93. https://doi.org/10.1111/j.1467-9833.2010.01486.x.

Rosen, Gideon. "Culpability and Ignorance." *Proceedings of the Aristotelian Society* 103 (2003): 61–84.

Ryan, Meghan J. "Compensation for Wrongful Convictions in the United States." In *Compensation for Wrongful Convictions*, edited by Wojciech Jasiński and Karolina Kremens, 161–82. Routledge, 2023.

Simms, Tina. "Statutory Compensation for the Wrongly Imprisoned." *Social Work* 61, no. 2 (April 1, 2016): 155–62. https://doi.org/10.1093/sw/sww003.

Slifer, Stephanie. "How the Wrongfully Convicted Are Compensated for Years Lost." CBS News, March 27, 2014. https://www.cbsnews.com/news/how-the-wrongfully-convicted-are-compensated/.

Thomson, Judith Jarvis. "Rights and Compensation." *Noûs* 14, no. 1 (1980): 3–15. https://doi.org/10.2307/2214886.

Valentini, Laura. "Justice, Disagreement and Democracy." *British Journal of Political Science* 43, no. 1 (January 2013): 177–99. https://doi.org/10.1017/S0007123412000294.

Viehoff, Daniel. "Legitimacy as a Right to Err." *Nomos* 61 (2019): 174–200.

Wellman, C. H. "The Space between Justice and Legitimacy." *Journal of Political Philosophy* 31, no. 1 (2023): 3–23. https://doi.org/10.1111/jopp.12272.

Wellman, Christopher Heath. "No One Defends the Status Quo." In *Rights Forfeiture and Punishment*, edited by Christopher Wellman, 172–92. Oxford University Press, 2017. https://doi.org/10.1093/acprof:oso/9780190274764.003.0008.

Yost, Benjamin S. "Death and Retribution." In *Against Capital Punishment*, 32–72. Oxford University Press, 2019.

Zalman, Marvin, and Julia Carrano. *Wrongful Conviction and Criminal Justice Reform: Making Justice*. Routledge, 2013.

Conclusion

This thesis considered a variety of philosophical problems in light of different elements of criminal legal systems. Chapter 1 demonstrated that the justification of the practice of arrest, despite its harms, has not been sufficiently challenged or question. Chapter 2 argued for calling pre-trial detention what it is—pre-trial punishment—which helps clarify its impermissibility. Chapter 3 identified numerous harms that occur when testimonial injustice occurs at trial. Chapter 4 contended that defense-introduced statistical evidence can help ensure proper convictions and reduce the likelihood of wrongful convictions. Chapter 5 demonstrated that, even if a person can deserve death, the practice of capital punishment is inherently and inescapably disproportional. Finally, Chapter 6 showed that compensation for over-punishment is more defensible than compensation for under-punishment.

These harms have likely received insufficient popular, political, and scholarly attention in part because the harms impact those who have been muted by society. Brian's term, "muted minority," suggests that people who are incarcerated experience hermeneutical marginalization. Hermeneutical marginalization occurs when a group is socially disempowered and thus systematically excluded from professions that shape societal vocabulary and concepts (e.g., academia, journalism, and politics).[343]

Hermeneutical marginalization has significant ramifications not only for those who are marginalized but also for society more broadly. When important social concepts, definitions, and justifications are under-developed, unjust harms are ignored or overlooked. Taken together, the essays of this thesis aimed to illuminate some of these harms in the hopes of calling more attention to the experiences of the muted minority. In so doing, they aim to argue for the mitigation of ongoing harms without losing sight of the need for a structural overhaul of criminal legal systems.

Bibliography

Fricker, Miranda. "Hermeneutical Injustice." In *Epistemic Injustice: Power and the Ethics of Knowing*, edited by Miranda Fricker, 147–75. Oxford University Press, 2007. https://doi.org/10.1093/acprof:oso/9780198237907.003.0008.

[343] Fricker, "Hermeneutical Injustice," 153.