

The Language of Punishment: Reluctant Bureaucrats, Public Morality, and the Law*

Isaac Dalke

Brendon McConnell

Keywords— Bureaucracy, Public Morality, Punishment, Text-as-Data, Causal Inference

*Thank you to Christopher Muller, David Harding, and Tyler Leeds for feedback on earlier versions of the paper.
Corresponding Author: Isaac Dalke, Post-Doctoral Fellow, Network Science Institute, Northeastern University and Departmental Fellow, Sociology Department, Harvard University. Mailing address: 177 Huntington Ave. Floor 10, Boston MA 02115. Email address: idalke@fas.harvard.edu; Non-corresponding author: Brendon McConnell, Associate Professor, Department of Economics, City St George's, University of London. Email address: brendon.mcconnell@gmail.com.

Abstract

A long tradition in sociology takes the law to encode public morality. Yet we know little about how it is given shape in administrative processes. How do street level bureaucrats exercise discretion in not only applying the letter of the law, but voicing its moral current? We examine a natural experiment in California where voters passed a law that pushed prison administrators to be more condemnatory in their decisions about who to release from prison. Specifically, we combine text-as-data methods with a regression discontinuity design to analyze 11,704 parole hearing transcripts. We find that parole decisions became simultaneously more punitive in their outcomes but less morally expressive in their substance, increasing incarceration while draining it of moral content. The findings reveal the ambivalent relationship between the broader emotive dimensions of punishment and those who enact it, while providing a novel window into how bureaucrats explain and justify their decisions.

1 Introduction

Bureaucracies are classically machines of rationalization: They produce decisions, and those decisions are supposed to be guided and constrained by reason (Weber, 1946). Yet however tight the rules, bureaucratic actors retain discretion in how to apply them (Lipsky, 2010). A key way that scholars have sought to understand the exercise of discretion focuses our attention on the personal moral dispositions of “street-level” decision-makers and how they apply their own normative beliefs to navigate formal rules (Musheno and Maynard-Moody, 2003; Maynard-Moody and Portillo, 2010; Zacka, 2017; Shiff, 2021; McDonnell, 2025). That is, in this traditional approach scholars examine how bureaucrats creatively engage formal decision-making frameworks to realize their personal preferences.

Typically, scholars of front-line bureaucrats treat the rules of bureaucratic procedure as neutral and inert. Yet lurking in the background is the fact that the rules themselves often explicitly claim to channel and represent *public* moral sentiment. Indeed, the symbolic power of those who act in the name of “the state” or “the public” rests in their claims to represent the general interest (Bourdieu, 1994, 2014). When applying the rules, decision-makers are giving shape to this distinctive public morality on which their authority rests. How do frontline bureaucrats channel or deflect such public morality in their routine decision-making?

Perhaps nowhere is this more true than in the criminal-legal system, where decisions about punishment explicitly express moral sentiments about right and wrong (Durkheim, 2013; Smith, 2008). The criminal legal system is full of pronouncements by *public* servants such as district attorneys, judges, and police officers about the *public* interest in detaining, incarcerating, and otherwise meting out punishment to individuals. We follow Mills (1940) in treating such explanations of action as loosely coupled to the action itself. In this paper, we ask about the relationship between the moral vocabulary of punishment and the decision to punish in a specific high-stakes bureaucratic setting: hearings about whether to release someone from prison.

We investigate this question in the context of California. The state has been on the forefront of “law and order” penal populism over the last several decades, or the shift towards moral outrage and condemnation as a political driver of penal policy (Pratt, 2007; Aviram, 2020). Specifically, we investigate a 2008 voter initiative known as “Marsy’s Law” that sought to push decisions about who to release from prison in a more punitive direction. In California, the parole board oversees release decisions for all life-sentenced prisoners with an indeterminate sentence, roughly a quarter of the state’s prison population. In each hearing the commissioners decide whether to release the person up for parole, and if they deny release,

they set a “denial length” which sets when the person will be eligible for another hearing. The new law increased parole denial lengths, set the default option at the maximum denial length, and symbolically re-affirmed the importance of the victim in parole board proceedings. How did the punitive moral currents encoded in the law change commissioners’ decisions, and the moral tenor of those decisions?

Such a question is traditionally difficult to answer with classical methods. We gain new empirical traction with a novel combination of a causal inference-based research design coupled with Natural Language Processing (NLP) tools that allow us to extract nuanced information about the content, tone, and moral charge of parole commissioners’ speech. Our approach allows us to provide unique evidence on the relationship between the severity of outcomes and the stated reasons for those outcomes, and how both change under shifting legal conditions. We analyze 11,704 parole hearing transcripts for hearings held between January 1, 2007 and December 31, 2010.¹ We first use Natural Language Processing (NLP) methods to extract features of moral reasoning from commissioners’ decisions based on the popular Moral Foundations Theory (Haidt, 2012; Graham et al., 2013) along with other important signatures of commissioner speech. We then use a Regression Discontinuity (RD) design to assess how those features change after implementation of the new Law.

We provide evidence that the commissioners adapted to the Law with reticence and caution. The Law had no effect on the number of people granted parole, and doubled the average denial length. In terms of their reasoning, commissioners took longer deliberating and yet provided shorter explanations of their decisions. They also mentioned the new Law in a quarter of their decisions, and leaned into official documentation and information. As a whole, their decisions increasingly emphasized fairness and authority, but they relied less on all “moral pillars” (Haidt, 2012) in their explanations. In other words, their decisions became less morally charged and more procedural.

We further dis-aggregate results by race of the person up for parole. Surprisingly, against a backdrop of disparate racial impacts across the criminal legal system, we find no differences in the law’s impact on grants between white and black people up for parole, while we find some evidence of longer denial lengths for Hispanic people. However, we find that commissioners deliberated for longer in cases concerning white prisoners, and they reached for different moral vocabularies after the law to justify their decisions for different racial groups. In particular, hearings for black people up for parole tended to involve greater emphasis on authority and loyalty after the law, while commissioners expressed more emphasis on degradation and harm in hearings for Hispanic people.

Put together, the new law led commissioners to be much more severe in their decisions but also

significantly more guarded and reserved in their moral reason-giving. Ironically, the impact of a law designed to make hearings more morally expressive resulted in making decisions appear more bureaucratic at the same time that they became more punitive.

The findings contain implications for scholars of morality, bureaucracy, and punishment, while advancing the methodological integration of text-as-data tools with causal inference. For scholars of frontline bureaucrats, the findings demonstrate that statements of public morality are not linked to underlying decisions in a mechanical or simplistic way. How street level administrators give shape to broader moral currents in their decisions is an essential dimension of bureaucratic discretion that is often treated as secondary to the outcomes of decisions and subsumed under personal morality. Yet the moral justification of decision-making is a distinct part of public administrative processes and bureaucratic authority in its own right.

The consequences of this symbolic dimension are magnified in the realm of punishment. For scholars of the penal state, we build on scholarship that examines how practices of punishment interact with popular punitive sentiments. Over the course of the “punitive turn” in American social policy, scholars have remarked on the rise of penal populism in driving harsher outcomes, in tandem with increased managerialism in the exercise of punishment (Feeley and Simon, 1992; Pratt, 2007; Wacquant, 2009). The standard explanation for the rise of managerialism is that the unprecedented growth of people swept into the criminal-legal system overwhelmed administrators and increasingly directed their attention towards purely technocratic matters of controlling criminalized populations (Feeley and Simon, 1992). Our findings suggest an alternate, complimentary account – the possibility of a feedback loop where public punitive impulses translated into harsher but increasingly bureaucratic practices of punishment, draining them of moral meaning while leading to calls for ever harsher measures.

Lastly, on the methodological front, our work helps to push forward the integration of NLP methods with causal inference frameworks (Ash and Hansen, 2023; Feder et al., 2022; Gentzkow et al., 2019) to answer central sociological questions that historically have been difficult to investigate empirically. In recent years, sociologists have fiercely debated the relationship between what people say and what they do (Jerolmack and Khan, 2014; Lamont and Swidler, 2014; Tavory, 2020; Small and Cook, 2023). Combining text-as-data tools with causal inference frameworks allow for a new and precise way to investigate the link between speech and action. In the organizational realm specifically, such designs offer a new way to approach situations where decision-makers must explain and generate reasons for their decisions, from criminal court cases to child custody proceedings to college admissions.

The paper is structured as follows. We begin by considering the role of the law in the daily decision-making of street-level bureaucrats, and what is to be gained by integrating an understanding of public morality. We then consider what makes the administrative practice of punishment a promising site for investigating the link between decision-making and public morality, before presenting the specifics of our case and empirical expectations. After that, we move into the data, research design, and empirical findings, before concluding with further implications for the study of morality, punishment, and administrative decision-making.

1.1 Street Level Bureaucracy, Discretion, and Moral Reasoning

Bureaucracies are organizations ruled by rules, in Weber's famous formulation (1946, page 196). Recently scholars have taken an interest in how the rules interact with moral frameworks held by frontline bureaucrats (Zacka, 2017; Shiff, 2024; McDonnell, 2025). In this section, we consider the way that a particular view of "the rules" as the detached, formal, neutral terrain of decision-making has shaped the study of frontline bureaucrats in a way that obscures the moral content of the rules themselves and their frequent claim to encode public moral sentiment. This lacuna is important precisely because, we argue, claims to act on public morality are a crucial foundation of bureaucratic authority.

Studies of bureaucratic decision-making processes are often grouped under the rubric of "street-level bureaucracy" as the place where rules meet people (Lipsky, 2010). This tradition explicitly counters views of bureaucratic administration as purely a top-down exercise, instead emphasizing how the interests, concerns, and dilemmas of those tasked with "doing bureaucracy," in the words of McDonnell (2020, page 10). The term "street-level," in its original formulation, captures a wide range of actors who routinely interface with members of a general public in the name of the state: judges, healthcare workers, code inspectors, and tax collectors all fall under the general umbrella of individuals who directly regulate access to state goods and impose state sanctions (Lipsky, 2010).

Such actors famously exercise a large degree of discretion and agency in how they interpret and apply the rules to individual cases. Such discretion, many argue, is inescapable (Lipsky, 2010; Zacka, 2017). Indeed, attempts to contain discretion through the creation of new rules often backfires or produces unexpected results, as when attempts to increase diversity in organizational hiring processes produces less diversity in new hires (Dobbin et al., 2015). New rules can expand or narrow the scope of actors' autonomy, but even in the most extreme cases – such as the adoption of algorithmic decision-making – discretion is not dispensed with but rather displaced to other decision-making points such as how to record

particular features of cases for the algorithm (Brayne and Christin, 2021). The ubiquity of discretion animates a series of normative concerns about whether the rules are applied consistently across cases, both by the same bureaucrat and across bureaucrats. And ample evidence shows how discretion can result in unequal access and treatment across groups on the basis of race, class, and gender (Schram et al., 2009; Maynard-Moody and Musheno, 2012; Kutateladze et al., 2014; White et al., 2015).

These concerns have given rise to a wide-ranging research program that looks at how bureaucrats wield their discretion (Maynard-Moody and Portillo, 2010; Portillo and Rudes, 2014). A key move in this literature is to distinguish between bureaucrats' moral sensibilities and the formal law, and look at how the two interact. For example, Shiff (2021) writes, "in the course of applying rules to complex cases, frontline actors negotiate a complex interplay between a claim's formal eligibility for a prescribed right/benefit under codified law and its perceived deservingness for that right/benefit as shaped by shared moral schemas" (page 339; italics in the original). This is echoed in Musheno and Maynard-Moody (2003), who argue that bureaucrats experience their work not as a formal and detached application of rules to cases, but rather as the process of aligning the rules to their own judgments of deservingness. In a similar spirit, Zacka (2017) argues that moral dispositions are key to understanding the irreducible nature of bureaucratic discretion: personal moral dispositions precede and guide applications of rules to cases (pages 90-95). Meanwhile, Watkins-Hayes (2009) shows how these sensibilities are deeply shaped by the professional and social identities of bureaucrats – what a bureaucrat thinks others deserve depends on how a bureaucrat thinks about herself.

In this formulation, the law hits the bureaucratic ground floor pre-formulated and packaged as a set of formal rules for decision-making. In turn, bureaucrats creatively engage and interpret these rules through their individual dispositions. The law can clash with individual bureaucratic decision-makers' personal moral schemas in ways that generate deep dilemmas and spur creative engagement to align rules to schemas. This "discordance" (Shiff, 2021) is essential to understanding the embodied, concrete decision-making of frontline state actors.

Yet what this formulation misses – counterposing morality to the law – is that the law itself codifies a form of *public* morality. That is, the law is not simply a value-neutral set of rules, but rather supplies its own superordinate set of values that claim to be grounded in the public good (Bourdieu, 1994, 2014). In a basic sense, bureaucracy is an organizational design that aims to subsume individual moral dispositions to this public morality (McDonnell, 2020). It specifies not only the means at hand for bureaucratic actors, but also the aims to which they should be oriented. If we think of claims to public morality as

constituting an important dimension of bureaucratic authority, then it also is an important part of the discretion that frontline bureaucrats exercise. From this different angle, bureaucratic discretion involves how individual bureaucrats navigate not only the law as such, but also the public morality it encodes. As many note, giving reasons for decisions constitutes an important part of the work of the state's frontline bureaucrats, above and beyond making decisions (Zacka, 2017; McDonnell, 2020, e.g.). The importance of this public morality to frontline bureaucracies can be further seen through those who face their consequences: citizens and petitioners who encounter public bureaucrats experience not individual moral dispositions, but rather some larger entity called "the state," and through those interactions come to know what constitutes its claims on and obligations to them (Soss, 1999; Auyero, 2012; Lerman and Weaver, 2014).

A crucial dimension of how decision-makers wield their authority, then, is how they give voice to the collective sentiments that are braided into the law. What makes decisions by *public* officials more scrutinized, more conflicted, and more symbolically consequential, is precisely their grounding in claims to represent this public morality. And regardless of personal disposition, frontline decision-makers are deputized to give voice to such a public morality. Nowhere is this more true than in criminal law, which we turn to next.

1.2 Public Morality and Punishment

In this section we turn to Durkheim (2013) and his followers to understand how the law captures and claims to represent collective moral sentiment, particularly in the realm of punishment. However, the way that public moral currents translate into administrative action within the state is by no means direct. We develop a framework based on Mills (1940) to approach statements of public morality as a form of action in itself, and use it as a launching point for understanding how administrators give voice to public morality.

Since Durkheim, scholars have observed that symbolic condemnation is central to punishment (Durkheim, 2013; Smith, 2008; Garland, 1990; Wacquant, 2009). Through this symbolic dimension, those taking inspiration from Durkheim argue that punishment draws boundaries around and generates hierarchies between in-groups and out-groups, affirms the "collective conscience," and gives expression to outrage over violation of "the sacred" (Smith, 2008, pages 13-14). For Durkheim (2013, pages 67-68) himself, the law's "interpreters of its collective sentiment" act in a mechanical and unambiguous way to realize public morality in individual acts of punishment. But such collective sentiments are, of course, not

always universally shared, nor is their application a simple process. As Ferguson (2021) shows, different actors within the law can hold strongly divergent ideas about the nature of the collective sentiment they represent. And Garland (1990, pages 50-58) calls the modern legal structure of punishment a “compromise formation” that is the outcome of a mixture of public moral sentiment, political contestation, and bureaucratic concern. Yet, he argues, “penal codes do not radically change public sentiments so much as impose a particular ordering and organization upon them” (pages 57-58). This orientation to punishment – bringing Durkheim’s concerns about collective moral sentiment to Weber’s diagnosis of public administration – suggests that attempts to define public morality is at the heart of penal politics. The bureaucratic administration of punishment, in turn, involves rituals that bring such public morality encoded in law into contact with specific people and cases.

The symbolic dimension of punishment is socially consequential in itself. Statements of public morality express and affirm distinctions between good and bad or deserving and undeserving actions and people, and they pervade the criminal-legal system from courtrooms to police stops to parole hearings. Scholars have conceptualized criminal-legal processes as “degradation ceremonies” (Garfinkel, 1956) that impart a new, symbolically defiled status on the individual accused of legal and moral transgression (Herbert, 2022; Van Cleve, 2016). A large body of scholarship investigates these assertions of public morality to explicate their underlying symbolic structure and their relation to social hierarchies of race, class, and gender (Van Cleve, 2016; Richie, 2012; Gustafson, 2013; Carlson, 2018; Rios et al., 2020). In turn, “criminality” is itself a powerful symbolic principle of vision and division, distinguishing who deserves what across public domains, from medical care (Lara-Millán, 2014) to education (Hirschfield and Celinska, 2011) to immigration (Hagan et al., 2008). Scholars have proposed that, through this symbolic dimension, the criminal-legal system has become a “race-making” institution: through differential targeting, punishment separates social groups as racially distinct and marks some as symbolically tarnished (Wacquant, 2009; Muhammad, 2010).

Yet studies of the symbolic dimension of punishment and its consequences are often detached from the specific adjudication and administration of punishment: the stigmatizing effects and degrading rituals are treated as independent of the degree of punitiveness and nature of the sanction imposed. This is a lacuna reaching back to Durkheim’s original formulation of the law’s moral contents (Garland, 1990, Chapter 2). In this paper, we ask about the specific relationship between how bureaucratic actors voice the public moral sentiment of punishment, and the punitiveness of the decisions they reach. To answer, we need to more specifically formulate the nature of such statements of public morality.

Formal statements of public morality — judges pronouncing from the bench, district attorneys presenting at a news conference, parole boards announcing whether to release someone from prison — are often highly staged. Rather than unreflexive utterances or transparent reflections of the speaker’s inner feeling, we would do well to view them as attempts to use the available vocabularies of public morality to explain and justify one’s decisions to wield state authority in a particular way. Long ago, C. Wright Mills (1940) proposed to “approach linguistic behavior, not by referring it to private states in individuals, but by observing its social function of coordinating diverse actions” (page 904). For Mills, this social function is to justify one’s past and future actions through appeal to widely accepted and acceptable “vocabularies of motive,” which are provided by the concrete social situation (pages 907-908). For Mills, a person’s action, the vocabularies for describing and justifying that action, and the social context of action exist in a complex interplay.

This approach has two central implications for understanding public morality and punishment. First, Mills invites us to view reason-giving as only loosely coupled to the act to which it refers. This opens up empirical questions about the relationship between a decision and the reasons given for a decision. In a related but distinct second point, this relationship is mediated by social context: reason-giving as such involves matching a past or intended action to reasons that are plausibly acceptable within the specific social situation. This second point suggests another series of questions about the fit between reason-giving and social context. Such a theoretical approach – first, separating action from justification, and second, placing justification within social context – is in line with much recent social psychology research on moral reasoning, which finds that people reach for shared moral vocabularies to justify their initial, instinctual intuitions about the best decision, rather than beginning with a formal moral framework from which they then derive decisions (Haidt, 2012; Jerolmack and Khan, 2014; Mercier and Sperber, 2017).

What are the implications for bureaucratic expressions of public morality and decisions to punish? In sum, this approach argues that moral reasoning in social contexts is about finding the words that align with the decision one has reached, and it invites empirical investigation into the specific ways that such alignment happens in practice. It is important to note that this alignment is distinct from how personal morality syncs to the law: the institutional design of law does not simply constrain bureaucratic actors; it provides the (literal) terms on which they act. However, expressions of public morality are related to, but distinct from, the contents of punishment itself: In the Durkheimian tradition, the contents of punishment are justified by the moral dimensions of punishment, but are not directly derived from

it. Adding another wrinkle, administrative decision-making is not mechanically descended from public morality. In a highly disjointed and decentralized criminal-legal system, where most decision-makers are only indirectly accountable to the public, they are not giving voice to “the mob” but rather are situated within a bureaucratic apparatus that is defined by explicit rules about both the contents of punishment and the acceptable reasons for enacting punishment.

This paper poses and investigates two questions that follow: First, what is the relationship between public moral sentiment and administrative expressions of public morality in the criminal-legal system? Second, how do these statements of public morality correspond to the severity of punishment? These questions push to further situate the symbolic dimensions of punishment within both the broader context that criminal-legal officials speak, and the specific decisions that they make. To gain leverage on these questions, we turn to a natural experiment to observe how decision-makers respond when the public signals desire for more severe and condemnatory punishment.

1.3 The Case: Marsy’s Law and the Parole Board

Parole boards are an ideal site for examining the relationship between administrative decision-making and public morality. Hearings to determine whether to release someone from prison, particularly in California, are grounded in deep moral reasoning about culpability, remorse, and personal reformation (Aviram, 2020; Dalke, 2024; Herbert, 2022; Shammas, 2019). As such, they play a central role in dramatizing and claiming to embody moral public sentiment.

In California, hearings are overwhelmingly for people serving indeterminate life-term sentences for morally-charged crimes of manslaughter or murder.² In 2010 people serving parole-eligible life sentences accounted for about 20 percent of the state’s prison population, or over 32,000 people in total (Weisberg et al., 2011, page 3). Life-sentenced prisoners are more likely to be men and tend to be older than the prison population as a whole, though they mirror the racial makeup of California’s prisons (Weisberg et al., 2011, page 16). The hearings are exceptionally high stakes: For this group, the only way out of prison goes through the parole board.

The state’s Board of Parole Hearings (BPH) oversees hearings, which typically take place in the prison where the parole candidate is incarcerated. These hearings are less formal than court procedures, and individual hearings can stretch for hours as commissioners question imprisoned people seeking release about their upbringing, the crime they are charged with, their time in prison, and their future aspirations. High profile cases, such as members of the Manson Family (Aviram, 2020) or the Menendez brothers

(Arango, 2025), attract wide media coverage and provide a stage for broader public discussion about the nature of harm, guilt, redemption, and punishment. The average parole hearing does not receive such scrutiny, but decisions are still infused with moral reasoning. Commissioners must legally provide “some evidence” for their decisions, which can range from a person’s current mental state to their disciplinary history to their lack of parole plans (Wattley, 2013, page 2).³ Yet most hearings hinge upon whether a person has taken “full responsibility” for the crime, expressed sufficient remorse, and demonstrated “insight” into what caused them to commit it, which puts wide discretion into the hands of commissioners (Aviram, 2020; Shamma, 2019; Dalke, 2024).

We analyze how commissioners responded to Marsy’s Law, or Proposition 9. Voters passed the ballot initiative on November 4, 2008 with 53.8 percent of the vote, and it went into effect on December 15, 2008.⁴ Aviram (2020, page 53) writes that the Law “is emblematic of California’s political culture and of the power of the victims’ rights movement,” and that it was specifically “designed to significantly influence the [parole] hearing process.” As a highly punitive, self-styled “victim’s bill of rights,” the Law sought to morally affirm the harm to the victim and the guilt of the offender in parole hearings (and at several other points in the criminal-legal system). It expressed this through increasing victims’ and family members’ rights to participate in hearings, explicit instructions on how to consider victims in decisions, and tripling the minimum and maximum denial lengths available for commissioners to impose. (It is worth noting that the Law did *not* change the underlying factors that commissioners can legally consider when arriving at a decision.) The Law’s punitive current was reflected in the broader public ballot campaign. The Law was named after Marsy Nicholas, a white college student who was murdered by her ex-boyfriend. Marsy’s brother, a billionaire corporate executive of a semi-conductor company, spearheaded the initiative and provided financial backing, while a coalition of victims’ rights groups was visible in the media throughout election season (Aviram, 2020, pages 52-54).

Marsy’s Law provides a unique case to investigate how penal administrators respond to signals of the public’s appetite for harsher punishment. In all of the ways outlined above, the Law pushed commissioners to more strongly condemn people up for parole and to emphasize the extent of the harm they caused. Marsy’s Law follows in the tradition of “penal populism” in California and the U.S. more broadly, where increasingly punitive penal policy is justified not on the basis of expertise or administrative concerns, but on claims about the popular will (Pratt, 2007). The relationship between penal populism and penal policy has been thoroughly debated and dissected (Beckett, 1999; Weaver, 2007; Wacquant, 2009; Foreman Jr., 2017). Yet we know relatively little about how penal administrators channel or deflect such currents of

public morality in their concrete decisions.

1.4 Empirical Expectations

Implementation of Marsy’s Law could have played out in a number of ways. We establish two measuring posts to gauge and contextualize our empirical findings about how the commissioners reacted in practice to the constraints and directives of the new Law. We call these measuring posts “bureaucratic resistance” and “bureaucratic embrace.”

On the one hand, commissioners could have resisted the push to become more punitive, both materially and symbolically. This reaction might be anticipated by frameworks of bureaucratic decision-making that give full primacy to individual morality or personal preference. If a commissioner thought that a one year denial was the right outcome for a case prior to the new Law, it is unclear why their personal moral calculus would transform afterwards. Instead, we would expect to see attempts to match the new denial choices to each individual’s prior moral preferences. Such effects might be predicted by the observation that externally-imposed attempts to constrain or shift the discretion of decision-makers often backfire (Dobbin et al., 2015). Further, we may even expect that clashing public mandate and personal disposition would lead to stronger articulations for decisions that cut against the moral tone of the new Law. In the extreme case, commissioner behavior would align with “sabotage” (Gates and Brehm, 2010), “guerrilla government” (Hollibaugh Jr. et al., 2020), or “extraordinary discordance” (Shiff, 2021), where decision-makers actively work to undermine directives. In its ideal-typical form, we label this possibility “bureaucratic resistance.”

In the opposite direction, commissioners could have embraced the changes. Think here of Zimbardo’s infamous prison guards who seized upon their newfound authority to punish with zeal (Haney et al., 1973). Even if commissioners’ fervor was unlikely to reach such heights, the new law contained many big and small signals designed to encourage any punitive impulses. A voluminous body of psychology research shows how decision “choice architecture” (Johnson et al., 2012; Thaler et al., 2013), such as the ordering of options and the addition or subtraction of options, can fundamentally shape a person’s ultimate decision. Two of the most well-known effects involve framing and anchoring. Framing refers to how a problem is presented to emphasize risks or opportunities, which can significantly shift the decisions individuals make (Kahneman and Tversky, 1984). Likewise, the default choice can “anchor” decisions, such that different defaults lead to different outcomes (Tversky and Kahneman, 1974). Marsy’s Law changed the choice architecture and foregrounded aspects of the hearing that were meant to push commissioners in a

much more punitive direction. Behavioral economic theories predict that such changes would “nudge,” if not shove, commissioners to be harsher. In such a scenario, we would expect denial lengths to at least triple in length (in accordance with the Law) and for the moral tone of the decisions to become more condemnatory. We call this ideal-typical possibility “bureaucratic embrace.”

To preview the findings: Rather than resisting or embracing, what we observe instead is that the severity of decisions and the tone of moral condemnation move in opposite directions. After Marsy’s Law went into effect we find that denial lengths became much more severe, more than doubling in length. This was less than the “embracing” scenario would expect, yet significantly longer than what would be predicted by the “resisting” scenario. However, we find that this increased punitiveness was accompanied by shorter, more formalistic, and more emotionally muted explanations. In other words, commissioners put the punitive changes of the Law (mostly) into practice without giving it voice, instead leaning into the rote administrative aspects of the process to describe their decisions. This suggests a reluctant compliance – enacting the punitive letter of the Law but backing away from its spirit.

Such results underscore the value and importance of treating the symbolic dimensions of bureaucratic practice as distinct from their material correlates and empirically investigating both in tandem. In the conclusion, we develop the implications for the study of bureaucracy and punishment. Next, we discuss our data and analytic approach before presenting more detailed empirical findings.

2 Data and Descriptive Evidence

2.1 Parole Transcripts and Administrative Data

The paper draws from transcriptions of 11,704 hearings for parole-eligible prisoners between January 1, 2007 and December 31, 2010 that resulted in either a grant or a denial. The hearings have been previously transcribed from an audio recording by the Board of Parole Hearings and serve as the legal account of the hearing. They are available to the public upon request under California law.⁵ The transcripts include the proceedings portion of the hearing where the commissioners intensively question the prisoner. The transcripts also include the decision portion of the hearing where the commissioners announce the outcome of the hearing and provide reasons for that outcome. The transcripts do not include the off-the-record deliberation period between proceedings and decision where the commissioners privately arrive at a decision.

We apply NLP information extraction techniques to the proceedings section of the decision to identify

basic characteristics of the hearing. This includes the prison location, the names of the commissioners, whether the prisoner is present, whether the prisoner has legal representation, whether any victims or victim's next of kin are present, whether the hearing is conducted through a translator, and whether the hearing has additional observers present. Following prior research, we identify whether the commitment offense includes a conviction for murder or a sex offense (Young et al., 2016). We also capture basic characteristics of the hearing logistics: the time at the beginning of the hearing, the time when commissioners enter into deliberations, the time when commissioners come out of deliberations, and the time when the hearing ends. Because commissioners are inconsistent about announcing the time, we can only reconstruct the full hearing time for roughly half of the transcripts.

During the study period there was some legal uncertainty about the legal standard for parole decisions (Hempel, 2010), and some hearings were conducted on the order of a court. We identify whether the beginning of the hearing includes the terms 'court order' or 'court-ordered' and exclude these from our analysis. We also exclude hearings from the post-Marsy period where the denial length aligned with pre-Marsy's Law denial lengths (1 year, 2 years, or 4 years), indicating that the hearing was held under the old standard. These may be hearings that were ordered by a court or that were rescheduled from the pre-period to the post-period for administrative reasons.

We use the unique identifier number for each prisoner to link each hearing to administrative data from the California Department of Corrections and Rehabilitation (CDCR). This data includes the prisoner's admission date, their county of commitment, and their state of birth. It also includes each person's prison-assigned race and ethnicity (Goodman, 2008). CDCR treats "Hispanic" as a racial category and as mutually exclusive from other categories. For some prisoners CDCR did not provide race or ethnicity, citing safety and security concerns.

Lastly, we focus on hearings where the Presiding Commissioner had experience with hearings before and after the Law. This allows us to rule out the possibility that any observed changes are not a result of shifting commissioner composition.

The final full sample for quantitative analysis includes 5,623 hearings where (a) the person up for parole is categorized as Black, White, or Hispanic; (b) the hearing was not held under court orders; and (c) the presiding commissioner for the hearing held at least 5 hearings before Marsy's Law went into effect and at least 5 hearings after.

2.2 Using NLP to Measure Public Morality

For the main analysis, we focus only on the decision portion of the transcripts. There may be additional people who speak in the decision (for example, a lawyer or a prisoner may ask a clarifying question), so we used recurring textual markers to isolate commissioner speech. The final corpus includes 26,031,691 tokens in total, and the mean decision length is 2,224 tokens. We use a variety of methods to generate measures of the text to assess changes to commissioners' language, including counts, scores, and semantic axes based on a word embedding model.

Counts: First, we create counts of the number of unique words in the vocabulary of the decision, the overall number of tokens in the decision, and the number of sentences in the decision. We also generate basic counts of whether and how many times commissioners mention Marsy's Law, the rules and regulations governing parole hearings, and the words 'victim' or 'victims'.

Scores: We generate sentiment, certainty, and entropy scores. For sentiment, we use the valence dictionary from the NRC Valence, Arousal, and Dominance Lexicon (Mohammad, 2018). The valence dictionary provides ratings of 20,000 words assigned by human raters. We take the mean valence ratings over all of the scored words in a decision to get an overall valence score for each decision. For the certainty score, we rely on the Rocklage et al. (2023) certainty lexicon. The lexicon was developed using a series of text responses from online workers who were asked to express differing levels of (un)certainly. The researchers then identified terms associated with high and low levels of certainty, and underwent a subsequent human rating process to generate certainty scores for 5,103 n-grams. We average over the n-gram scores for each transcript to assign an overall score of how much certainty the language in a given decision projects. Last, we calculate the Shannon Entropy score for each decision. This measure captures how predictable the language in a decision is as a function of the probability distribution of all the words in the vocabulary.

Word Embeddings: Our analysis of moral language relies on word embeddings. We train a word2vec skipgram with negative sampling (SGNS) model (Mikolov et al., 2013; Goldberg and Levy, 2014) on all 11,704 decisions in the corpus. The SGNS model learns vector representations (or word embeddings) for each unique word in the vocabulary of the corpus. It does so by training a shallow neural network with a single hidden layer to predict the context around a word given the target word. (For example, if trained on the previous sentence the model may be given the target word 'neural' and trained to predict that the word 'network' appears in the immediate context around it.)

The coefficients from the hidden layer that the model learned to complete the prediction task are the

“word embeddings” that we can conceptualize as coordinates locating each word in a unique position within a shared “embedding space.” The dimensions of the embedding space are arbitrary on their own, but scholars have found relations between words in embedding space track meaningful relationships between words in ordinary understanding (Mikolov et al., 2013). At a high level, this is because the embeddings capture information about the contexts in which words appear, and much of a word’s meaning is encoded in its context. Scholars have used word embeddings to track how gender stereotypes, disease stigma, and the cultural meanings attached to class have changed over time (Best and Arseniev-Koehler, 2023; Garg et al., 2018; Kozlowski et al., 2019). In the criminal-legal context, scholars have used word embeddings to look at the adoption of economic language in the U.S. federal courts (Ash et al., 2018) and U.S. judge sentiment towards different social groups (Ash et al., 2022), among other applications.

Prior to feeding the text into the model, we convert all text to lower-case but otherwise engage in minimal pre-processing, as is recommended for word embedding models (Rodriguez and Spirling, 2022). We make a series of decisions about the model training parameters. The model only learns representations for words that appear at least 10 times in the corpus. We train a model with 300 dimensions and use a context window of 6 words, which is in line with convention (Rodriguez and Spirling, 2022) and past research on the California parole board transcripts (Dalke, 2024). Word embedding models are randomly initialized and sensitive to the composition of the corpus they are trained upon. Both introduce an element of inherent randomness to the training process, meaning models will vary across different implementations even with the same corpus and model parameters. To account for this, we follow the bootstrapping procedure recommended by Antoniak and Mimno (2018). We train 100 models on different subsets of the corpus that have been sampled with replacement to obtain the original sample size. Throughout the paper we report findings that have been averaged over the 100 bootstrapped models.

While we restrict the sample for the final Regression Discontinuity analysis, we train the word embedding model on all decisions in order to capture the patterns of language use for the parole board as an institution, and because word embedding model performance tends to improve with corpus size. In training the word embedding model on the full corpus, we make the assumption that commissioners are not using words in new or novel contexts as a result of Marsy’s Law. We can empirically probe this assumption using word embedding models separately trained on the pre- and post-Law hearings and the procedure recommended by Rodman (2020) for assessing language change between two time periods. For the 5,000 most frequent words in our corpus the mean change score between the pre-Law and post-Law periods is 0.98 out of 1, with 1 representing no detectable change. This suggests that there

was no substantive shift in word usage across hearings following the implementation of the Law.

Semantic Axes: We track the moral valence of decisions in embedding space by locating each decision along a “semantic axis” between two opposing poles (such as “good” and “bad” or “man” and “woman”) (An et al., 2018). This requires us to generate a vector representation of the decision and a vector representation of the semantic axis using our embedding model.

To create the semantic axis we rely on the Moral Foundations Dictionary (MFD) (Frimer et al., 2019), as has become increasingly common in NLP research (Best and Arseniev-Koehler, 2023; Enke, 2020; Mendelsohn et al., 2020). The dictionary provides positive and negative words for each of five moral dimensions: sanctity/degradation, loyalty/betrayal, fairness/cheating, authority/subversion, and care/harm (Frimer et al., 2019). For any given dimension, we average over all of the vector representations of positive words in the MFD, then average over all of the negative words, and finally subtract the averaged positive vector from the averaged negative vector to obtain our semantic axis (An et al., 2018). Note that for a word to be included from the dictionary, the embedding model has to have learned a representation of that word. This means words that do not appear in the decisions are not included in the axes.

Next we create a vector representation of the decision text by averaging over the vector representations of every word in the text. We weight the words by inverse document frequency to give more weight to the words that are more distinctive in a given decision, on the assumption that they are more informative. We then locate the decision on the axis by calculating the cosine distance between the vector representation of the decision and the vector representation of the semantic axis. We center the distance measure at 0 and scale it to have a standard deviation of 1. We average over this standardized distance measure for all 100 models to get the final semantic axis measure we use in the analysis. We follow this procedure for all five moral foundation pillars. We run additional analyses looking at the cosine distance to each positive and negative pole separately, as well as a reduced model that uses only the 10 most frequent words from each pole to construct the axes.

We use the semantic axis framework to investigate several other features of language use in addition to the MFT pillars. We generate our own “blame” axis by compiling all synonyms of the word “blame” from common thesauruses to generate the positive end of the axis and all antonyms to generate the negative end of the axis. We also use the word embedding model to investigate whether Marsy’s Law resulted in any change to the extent that commissioners lean on themselves (using first person pronouns), official documentation (such as a risk assessment or work evaluation), or other authority figures (such as prison guards or psychologists) in their decisions. We do this to investigate whether there is any shifting

basis of authority in how commissioners justify their decisions, such as shifting from invoking their own judgment to the judgment of other authority figures or authoritative sources of information. First, we simply count how many times any first person singular or plural pronoun appears in the text of the decision. We also convert the list of pronouns into a semantic “pole” in embedding space and create a pole for documentation and authority figures. To generate the authority figures and official documents poles we start with a list of the most frequent nouns used in the hearings and identify terms that generically refer either to documents or to authority figures, respectively. We then reference thesauruses to add to the lists and query the word embedding model for the nearest neighbors to in embedding space to uncover terms that are used in similar contexts to the words in our initial list. With the two word lists compiled, we follow the same procedure to create an “official documents” pole and an “authority figure” pole. As it is not clear that pronouns, authority figures, or official documents are conceptually opposed to each other, we do not generate a set of semantic axes and instead report results only for the individual poles.

2.3 Initial Descriptive Evidence

We present summary statistics for both our full sample and our main working sample in Table 1. The working sample reflects a subset of 3,295 hearings that fall within the optimal date bandwidth for our Regression Discontinuity Design (for a fuller description see Section 3). This table highlights the absence of any meaningful sample selection issues when we restrict the full sample to our working sample.

[Table 1 here]

Before implementing our regression discontinuity approach, we provide initial evidence of the impact of Marsy’s Law on parole hearing outcomes. We collapse the data to the monthly level for these graphs, and start the month on the fifteenth, in order to align our results with the introduction of the Law. In Figure 1, we present results for the full sample period we have available – January 2007 until December 2010. In Appendix A we provide further visual evidence of how outcomes changed over the study period. We treat parole grants as having a denial length of 0. Parole denial lengths jumped discontinuously with the introduction of Marsy’s Law, from a baseline of 2.1 years prior to the Law, to 4.1 years after the Law came in to effect. The proportion of prisoners denied parole three years plus tripled, from a baseline of 28 percentage points to a post-Marsy’s Law average of 86 percentage points. A denial length of five years plus, which prior to the Law’s enactment occurred in 8 percent of hearings, occurred in every 3 of 7 cases after the Law – 43 percent of the time. Such summary statistics motivate our RD design.

[Figure 1 here]

Meanwhile, we can use the distribution of denial lengths to gain initial traction on whether commissioners made more or less harsh decisions than prior to the implementation of the law. We do so in the following section, and then present descriptive evidence of how the Law shaped the language that commissioners used to justify their decisions.

2.3.1 Marsy’s Law and Denial Lengths: Decomposition Exercises

We seek to gain further initial descriptive evidence on how Marsy’s Law impacted denial lengths through a decomposition exercise. The Law tripled the denial lengths available to commissioners, but how this impacted outcomes was mediated through commissioners’ response to the new available options. The two decompositions we present here create a bridge between the pre-Law period and post-Law changes by introducing an intermediate step that holds fixed one of two dimensions: the ordinal ranking of denial lengths, or the absolute value of denial lengths. By doing so, we simulate the outcomes we would expect if commissioners had either fully embraced or fully resisted the new Law (as explained in Section 1.4).

The first decomposition investigates the possibility of bureaucratic embrace. For an ideal-typical “embracing” scenario, the important element of the choice set is the relative severity of a given denial length within the available choices, rather than the absolute length of the denial. That is, if purely driven by the choice set, we would expect commissioners to apply their same relative preferences for shorter, medium, and longer denials to the new set of denial length options. This decomposition allows us to compare the observed denial lengths against this expectation.

The second decomposition that we present captures the response to the Law if commissioners had been “resisting” its directives. Such a commissioner would not respond to any of the Law’s signals to become more harsh. Rather, if they maintained the same absolute denial length preferences as before, they would choose the denial length in the post-Law period that most closely matched what they would have chosen in the pre-Law period.

The decomposition proceeds in three steps. First, we calculate the mean of denial lengths in the pre-Law period. This is based on the choices parole commissioners made based on the set $d_0 = \{0, 1, 2, 3, 4, 5\}$.

Second, we calculate the post-Law denial lengths under the counterfactual embracing and resisting scenarios. Such a counterfactual simulation step is in the spirit of a Kitagawa-Blinder-Oaxaca decomposition (Kitagawa, 1955; Blinder, 1973; Oaxaca, 1973) used in labor economics to decompose the source

of group-based differences such as gender wage gaps.

For the embracing scenario, we keep fixed the ordinal structure of the denial lengths commissioners chose in the pre-Law period, but we convert these to match the choice set in the post-Law period ($d_1 = \{0, 3, 5, 7, 10, 15\}$), and again calculate the mean denial length.⁶ This second step is a counterfactual simulation step – we simulate the denial lengths we *would* see if commissioners were to keep their pre-Law choice order fixed when faced with the post-Law choice set. For example, if a commissioner opted for the third lowest denial length in the pre-Law period (2 years), she would now opt for a 5 year denial (the third lowest denial length post-Law). This provides a lower bound for the “embracing” framework, as it does not take into account the specific anchoring or broader social signals that could push commissioners to hold *more severe* relative preferences after the law. Figure 2 presents this decomposition graphically, with the “embrace” scenario captured in orange. Note that as we move from Figure 2a to the orange bars in Figure 2b, the shape of the distribution is simply transposed onto the new denial length options.

For the resisting scenario, we project the post-Law outcomes as if commissioners chose the denial length choice from the post-Law choice set that is as close as possible to what they chose prior to the introduction of the Law. Table 3 outlines the mapping procedure we use for this counterfactual simulation. Three sets of choices are noteworthy here. For the choice of 4-year denials in the pre-Law period, we randomly allocate half of the decisions to 3 years and the other half to 5 years, as 4 years is no longer in the pre-Law choice set. Second, we allow for the fact that in the pre-Law period, commissioners choosing the maximum denial length of 5 years may have been constrained by the 5-year ceiling. Hence in the post-Law period, we allocate pre-Law denial length choices of 5 years to denial lengths of 5, 7, 10, and 15 years with equal probability. Third, we make the conservative assumption that commissioners would not issue more grants. This means that the decomposition presents an upper bound for the resisting scenario.

For the third step, we calculate the distribution of denial lengths in the post-Law period and compare against the simulated distributions. We present the results of this decomposition exercise in Table 2, and present the distributions of denial length across the three cases in Figure 2.

[Table 2 here]

In row 1, we present the pre-Marsy’s Law mean of denial length. In row 2, we show the two counterfactual possibilities along with the observed distribution of denial lengths after Marsy’s Law took hold. For the “embrace” scenario indicated by the orange bars, we find that the simulated counterfactual mean denial length would have been 5.3 years. This is in contrast to what we actually find in the post-Law period – a mean of 4.1 years. We implement a Pearson’s χ^2 test to compare the “embracing” and observed

distributions in the post-Law period, strongly rejecting equality of distributions. This provides initial evidence that commissioners offset the longer parole denials mandated by Marsy's Law by choosing lower denial lengths from the period options available to them.

[Figure 2 here]

[Table 3 here]

On the other hand, they also did not fully resist the Law's punitive impulses. The distribution of denial lengths under the "resistance" counterfactual is represented by the blue bars. Under this scenario, we would expect to see mostly 3-year denials, and very few denials longer than 5 years. The simulated distribution anticipates a mean denial length of 3.3 years, substantially lower than the actual denial length we observe (4.1 years). Once again, we use a Pearson's χ^2 test to compare the counterfactual and actual distributions in the post-Law period, which statistically confirms that the simulated "resistance" distribution is distinct from the observed distribution.

In summary, the observed decision lengths falls above what we would expect if the commissioners had resisted the law, but below what we would expect if they had fully embraced it. This provides initial evidence of commissioners' ambivalence about implementing the law: they do significantly increase the lengths of the denials they issue, but do not go as far as to simply apply their old relative preferences to the new lengths, let alone extend beyond to lean into the more punitive dimensions of the law.

2.3.2 Marsy's Law and Moral Vocabularies

Next, we present descriptive evidence on the moral vocabularies of commissioners. The descriptive evidence suggests that Marsy's Law scrambled the moral justifications of commissioners in more complicated ways than simply leaning into moral condemnation or becoming morally resistant.

Figures 3(a) through 3(p) compare our primary NLP measures by denial length and time period. Prior to the implementation of Marsy's Law, figure (a) shows that commissioners used more words to articulate their decisions when granting parole and denying parole for the longest available lengths; this pattern holds in figure (b) after implementation of the Law, with a pronounced increase in the number of words in decisions to grant parole.

In figure (c) we observe that, prior to the law, language in decisions leaned into authoritative information (the language of official reports and prison documentation) for grants and the longest denials. After the Law, however, figure (d) shows that commissioners gravitated in the direction of authoritative

language across 7-year, 10-year, and 15-year denials – no longer simply for the longest denials. This is our first major indication that the way commissioners gave voice to their decisions shifted in response to the law. Notably, it cuts against the new Law, which mandated that commissioners begin with a default denial length of 15 years and provide increasing evidence to justify shorter denial lengths.

[Figure 3 here]

With our measures of “blame” (figures (e) and (f)) and the five moral pillars (figures (g) through (p)), we find a consistent scrambling of the “tilt” of decisions after the Law goes into effect. Prior to Marsy’s Law we observe consistent and relatively uniform gradients – the language in decisions slants in a more positive direction for grants and short denials, and becomes more negative for longer denials. In other words, commissioners in this pre-period appear more condemnatory when issuing longer denials. This tight coupling of decisions to punish and vocabularies of condemnation are as one would expect. However, the patterns across all six measures look dramatically different after the Law goes into effect: language for the longest possible denial length becomes more negative, while the middle range of denial lengths appears to largely become less negative — though unevenly so. This descriptive evidence suggests that Marsy’s Law unsettled the moral vocabularies of the decisions, upending the relationship between punishment and condemnation, but the language of the hearings does not appear to tilt uniformly in a more harsh or more positive direction.

To this point, we have only provided a descriptive picture of the Law’s impacts on commissioners’ decisions. However, this evidence is purely correlational, and does not allow us to identify the specific causal effects of Marsy’s Law. For example, the comparison across time periods may capture over-time trends that are unrelated to the impacts of the new Law. In the following section, we introduce our analytic approach for gaining causal traction.

3 Analytical Approach: Regression Discontinuity Design

Given the sharp, exogenous implementation of Marsy’s Law, the natural methodological candidate to use in this setting is a sharp Regression Discontinuity (RD) design. The RD design enables us to estimate the local average treatment effect (LATE). In this case, ‘local’ refers to proximity to the cutoff (15 December 2008) of our running variable (hearing date). We use a sharp RD design because our treatment variable is the implementation of Marsy’s Law, and it shifts sharply at the cutoff – all parole hearings prior to 15 December 2008 were subject to the pre-Marsy’s Law standards; all hearings after were subject to

the mandates of the new law. This design isolates the immediate impact of the law on commissioners’ decisions and reasoning, allowing us to examine how commissioners responded when forced into much harsher decisions than they had made previously.

We favor the use of an RD design over the more simple pre/post comparison that we could do in this setting for two reasons. First, a pre/post design would capture not just the impact of the implementation of Marsy’s Law, but also any other short run trends in outcomes unrelated to the Law. The RD design is insulated against such a concern – the specification that we outline below is robust to differential trends on either side of the Law’s implementation date. Second, estimates from a pre/post design are vulnerable to being biased by unrelated shocks to the parole process that occur within the estimation window. The RD design is considerably more robust to such threats to identification – differential trends on both sides of the cutoff will capture any extraneous random shocks to the outcome, and the key parameter of interest is estimated from the discontinuity in outcomes occurring *right* at the date of the implementation of the Law. Taking another look at the graphical evidence that we provide in Figure 1 we see both (i) the immediacy of the change in parole setting behavior around the cutoff and (ii) no evidence of other shocks to parole setting that occur very close to the implementation of the Law. Based on these observations we conclude that an RD design is the best-suited empirical strategy to use in order to obtain a clean and causal estimate of the impact of Marsy’s Law on parole hearing behavior.

3.1 Empirical Specification

Our core empirical specification takes the form of an RD in time:

$$y_{it} = \alpha D_{it} + g^D(z_{it}) + X'_{it}\beta + \theta_c + \pi_p + \varepsilon_{it} \quad (1)$$

where our running variable, z_{it} , represents the hearing date, $D_i = \mathbb{1}[z_{it} > 15 \text{ December } 2008]$ is an indicator for Marsy’s Law, and $g^D(z_{it})$ is a function of z_{it} before and after the introduction of Marsy’s Law. In our baseline setting, $g^D(z_{it})$ takes the form of a polynomial of order 2 on either side of the cutoff. α is our parameter of interest, the LATE of the enactment of Marsy’s Law. The vector X_{it} includes a set of prisoner-level controls, including race indicators, most serious offense indicators, indicators for being born in the US, and specifically in California. In order to allow age and time served – our two continuous control variables – to enter non-parametrically into the regression, we additionally control for age and time served quintiles. θ_c and π_p respectively denote commissioner and prison fixed effects. The error

term is ε_{it} . We use Eicker-White standard errors throughout.

3.1.1 Identification

In order to interpret our RD estimates as the causal effect of the Law, we require two key conditions to hold. First, we require the time of parole hearings to be non-manipulable. This rules out commissioners or prisoners shifting their hearing dates based on the enactment of Marsy’s Law, and thus rules out selection bias in our setting. Secondly, we require the continuity assumption to hold, which is grounded in the potential outcomes framework (Neyman, 1923; Rubin, 1974). The continuity assumption states that the potential outcomes (Y_{it}^0, Y_{it}^1) vary continuously through the cutoff. We require this assumption to hold in order to ensure we are capturing only the effects of the discontinuity in treatment assignment at cutoff, and not other changes as well. If other variables were changing at the cutoff, our RDD would conflate these changes with the true treatment effect. In this sense, the continuity assumption rules out omitted variable bias at the cutoff.

We first provide evidence on the lack of manipulation of hearing dates, our running variable, in Figure 4. In addition to the graphical evidence, we also provide a statistical test of manipulation, using the procedure of Frandsen (2017). This procedure allows us to test for hearing date manipulation (our running variable) given that dates are discrete. In order to implement the test, we need to choose a parameter, $k \geq 0$, which serves as a test leniency parameter. The larger the value of k , the less power the test has to detect manipulation. We choose the strictest possible value, $k = 0$. The resulting p -value ($p=.508$) confirms what we glean from a visual inspection of Figure 4: there is no manipulation of parole hearing dates.

[Figure 4 here]

We next provide supportive evidence for the continuity assumption. This evidence is necessarily indirect, as we never observe potential outcomes. In Table 5, we present RD estimates for all of our control variables. We estimate an unconditional variant of Equation 1, where we sequentially consider each of our control variables as an outcome:

$$x_{it}^k = \alpha D_{it} + g^D(z_{it}) + \pi_p + \varepsilon_{it} \quad (2)$$

where x_{it}^k refers to control variable k from our vector of controls x_{it} . Given the temporal structure of commissioner visits to prison, we include prison fixed effects in all specifications. We do not detect

statistically significant discontinuities in any of these variables – observable determinants of parole outcomes – at the cutoff.

[Table 5 here]

Combining the evidence we present, we conclude that the implementation of Marsy’s Law satisfies the necessary assumptions to implement a credible RD design. In the next section, we proceed with our empirical strategy.

4 Regression Discontinuity Design Results

We now present our core RD results. We start first with parole outcomes. We then move to a series of analyses based on the language commissioners used to talk about those outcomes. The first set of analyses document the longer denial lengths that commissioners handed down after the enactment of Marsy’s Law, while the second seek to understand how parole commissioners explained and justified those longer parole denial lengths.

4.1 Marsy’s Law and Parole Outcomes

In Table 6 we present RD evidence of the impact of Marsy’s Law on parole outcomes. In order to maintain comparability of our findings across the various outcomes, we use a common bandwidth for all RD analysis. We follow the approach of Calonico et al. (2014) to calculate the optimal bandwidth for denial length – our key parole outcome variable – and then restrict the sample based on this optimal bandwidth of 295 days around the enactment of Marsy’s Law. The RD results are stark, and confirm the findings we document in the descriptive results (Figure 1).

The results in Column 1 highlight that there was no change in the parole grant rate around the enactment of Marsy’s Law. In Column 2, we document a 2.6 year increase in parole denial length – an increase of 130% over the baseline denial length prior to the enactment of the Law. In Column 3 and Column 4 we present results for indicators of parole denial lengths of 3+ and 5+ years, which respectively increased by 227% and 543% compared to the pre-Law baseline means.

[Table 6 here]

In order to better understand how parole commissioners navigate the new menu of parole denial length options mandated by Marsy’s Law, we present a period-specific normalized parole denial length in

Column 5. In this analysis we normalize denial length by the period-specific range of denial length options.⁷ This move serves to disentangle the effect of the increased denial length options from how commissioners chose among those options. The evidence we provide in Column 5 suggests that, on average, commissioners opted for a lower denial length from the available, period-specific options after the decision. This confirms the suggestive evidence we presented from the decomposition exercise in Section 2.3.1: commissioners handed down longer denial lengths post-Marsy’s Law, but they shifted downwards their position in the set of (period-specific) denial length options. Such a shift tempered the pure denial length-increasing effect of Marsy’s Law. In other words, commissioners deflected some of the punitive intent of the ballot initiative, but by no means actively resisted it.

4.2 Marsy’s Law and the Language of Justification

We now turn to consider what we can learn about commissioners’ responses to the mandates of Marsy’s Law from the parole hearing transcripts. In particular we are interested in how commissioners justify the longer parole denials they hand down post-Mary’s Law and the degree of moral condemnation in their decisions.

Section 1.3 provided divergent “measuring posts” to make sense of the decisions. Based on changes to the broader context and specific choice architecture, the “embracing” hypothetical response anticipated that commissioners would use more morally charged language and particularly emphasize the harm and degradation to victims. In contrast, the “resisting” hypothetical expected commissioners to continue to apply their old reasoning unchanged or become more positive in their language, as a way of combatting the negative moral charge of the new Law. We present RD estimates for a series of dimensions from the parole hearings in Table 7.

[Table 7 here]

What we learn from columns 1 and 2 is that the parole mandates of Marsy’s Law led to a 17% increase in deliberation length (column 2), while it did not detectably change the length of the hearing proceedings (column 1). Turning to the decisions, we find an increase in commissioner reticence to explain their decision – the word count (column 3), sentence count (column 4), and vocabulary size (column 5) of the decision all fell, by 7%, 14% and 4% respectively. At the same time, we see that in almost a quarter of hearings, commissioners mentioned Marsy’s Law (column 12). Altogether, after the Law went into effect

commissioners began taking longer to reach a decision while talking less about how they reached their decision and invoking Marsy's Law in a significant number of cases.

However, commissioners did not appear more uncertain or unpredictable in their language use: our measure of words conveying certainty does not change (column 7), nor do we see more entropy in their language use (column 6; entropy is a measure of predictability which we take as another proxy for the certainty and consistency of language use in the decision). The Law also does not change the number of times that commissioners invoked the legal code and regulations that govern parole decision-making (column 11). While commissioners were more reticent to explain their decisions, they did not voice more uncertainty in their decisions; and while they invoked Marsy's Law they did not also invoke the overall legal architecture any more or less frequently than they had before. This suggests caution, but not vocal uncertainty, in how the commissioners adapted to the new legal environment.

The rest of our analysis documents if, and how, the moral tone of the hearings shifted as commissioners began issuing on-average more punitive parole denial decisions. Our first look considers how discussion of the victim (column 13), blame-related language (column 14), and basic sentiment (column 15) change in response to the Law. Column 13 shows that commissioners referenced the words "victim" and "victims" more often after the Law. However, column 14 considers whether the vector representation of decisions moved towards the language of blame or blamelessness, and reveals no change. Finally, a count-based sentiment score (column 15) shows no change in the overall positive or negative tone of the hearings. The "embracing" scenario would expect discussion of the victim to increase as we observe, but also for commissioners to express darker sentiment and lean into the blameworthiness of the person up for parole, which the evidence does not support.

Next, we consider the extent to which commissioners may have attempted to shift the basis of their decision by invoking themselves less as active deciders and emphasizing others' judgment or "objective" evidence. We investigate this by calculating how the embedding representation of each decision moves relative to vector representations of first-person pronouns (column 8), authoritative sources of information (column 9), and authority figures (column 10). Commissioners' decisions do not change relative to self-referencing pronouns (such as "I" or "ourselves").⁸ However, we do find a shift relative to external reference points – column 9 shows that post-Law, commissioners' language shifts towards authoritative information and documentation (including, but limited to the words: records, evidence, facts, documentation, reports) and away from authority figures (captured by the words police, guard, psychiatrist, clinician and officer among others). Such a shift in reference points suggests a move

to emphasize the procedural evidence for their decisions in the aftermath of the introduction of Marsy's Law.

Taking these initial measures together, commissioners became more reticent in their decision-making once Marsy's Law went into effect. They took longer to reach a decision and then used less word to explain their decisions. They invoked Marsy's Law in a quarter of the hearings, and leaned more into official documentation and information in their decisions. Meanwhile, so far we find no evidence of the moral opprobrium that the "embracing" scenario would predict. In other words, commissioners do not appear to be channeling the public moral currents embedded in the Law, nor do they appear to be counteracting them. Rather, the hearings appear to become increasingly procedural and perfunctory. Next we provide a more granular look at the impact Marsy's Law had on moral reasoning in the decisions using the Moral Foundations Theory framework (Graham et al., 2013; Haidt, 2012).

4.3 The Moral Foundations of Decisions

In this section we gain a deeper understanding on how the law impacted the moral slant of the decisions. Moral Foundations Theory (MFT) presents five dimensions or pillars of moral reasoning: sanctity/degradation, loyalty/betrayal, fairness/cheating, authority/subversion, and care/harm (Graham et al., 2013).⁹ The "embracing" scenario would lead us to expect greater emphasis on harm and degradation, given the emphasis on the sanctity of the victim. Meanwhile, the "resistance" scenario approach would lead us to expect an increasingly positive moral valence, as commissioners would be consciously counteracting the Law's punitive thrust.

We measure the location of a decision's vector representation along each of the MFT pillars to observe how the moral slant of decisions changed after the Law went to effect. Positive values indicate a move towards more morally positive language, while negative values indicate a move towards morally condemning language. Row (a) of Table 8 shows the differential use of MFT-based concepts post-Marsy's Law. The findings point to commissioners emphasizing notions related to fairness (column 4) and authority (column 6) after the law went into effect. Meanwhile we observe no changes to other pillars of moral reasoning.

Table 8 presents three additional sets of analysis related to each measure: in row (b) we consider changes relative to only the positive pole for each pillar; in row (c) we consider changes relative to only the negative pole each pillar; and in row (d) we re-run the primary analysis but reconstruct each pillar using only the top-ten most frequent words for each pole.¹⁰ The estimates based on (b) and (c) inform us

on the underlying source of the treatment effects we estimate in (a). For example, is the post-Marsy’s Law rise in fairness-related speech that we document in row (a) of Column 4 driven by an increase in language positively related to fairness concepts, or rather due to a fall in language negatively related to fairness? The restricted measures, based only on the top ten words, serve as sensitivity analyses, and answer the question: if we cast a tighter net around the concepts of interest, do we still estimate a consistent pattern of findings?

[Table 8 here]

The additional analysis highlight two key points. First, the move in the direction of language about fairness and authority post-Marsy’s Law is driven by a move away from the negative poles of both concepts. Second, the null findings for the loyalty and sanctity pillars mask increased distance from both the positive and negative poles of each concept, which offset one another to yield an overall null effect of Marsy’s Law. The evidence that we present in panel (b) and (c) for both loyalty- and sanctity-based pillars points to commissioners’ overall declining use of such morally-charged language in their reason-giving, positive or negative. Put differently, the evidence in Table 8 suggests that commissioners rely on less moralizing language in general in their post-Law parole decisions. That is, the justifications for harsher punishments become drained of the public moral sentiment that led to their adoption.

4.4 Heterogeneity Analysis by Racial Group

In order to further probe commissioners’ responses to Marsy’s Law, we provide the results of a series of heterogeneity analysis by race of the person up for parole in Figure 5. We present the results from further heterogeneity analysis in Appendix Section A.6. This includes results disaggregated by (i) most serious offense in Figure A5 and (ii) by commissioner severity in Figure A6.

[Figure 5 here]

In addition to presenting a full set of RD estimates by race in Figure 5, we highlight a subset of these results in Table 9, where we also present *p*-values for tests of equality of RD estimates across specifications. For parole outcomes, there are (surprisingly) few differences. We find no differences in how the law impacted chances of being granted parole across racial groups. For denials, Hispanic prisoners typically experienced worse outcomes post-Marsy’s Law, but these differences are rarely statistically significantly different. The exception is for normalized denial lengths – although both white and Black prisoners

experienced lower normalized denial length post-Marsy’s Law, their Hispanic counterparts did not. This difference in RD estimates is statistically significantly different from zero at conventional levels of significance (see the p -values in the bottom three rows of Table 9).

We also find few differences across the core hearing details. Such a series of null findings underscore the value of our focus on moral reasoning in this work, as it is here where we find key differences across racial groups. In discussing the findings from our baseline estimates in Section 4.2, we noted that commissioners draw on fewer moral pillars when explaining their parole decisions post-Marsy’s Law, while placing greater emphasis on fairness and authority. In both Figure A5 and Table 9, we find differences in how the law impacted moral reasoning by race of the person up for parole. The RD estimates for Hispanics are negative for all 5 MFT pillars. Two of these estimates are statistically significantly different from zero (columns 3 and 6), and all five estimates are statistically significantly different from one of the other two race-specific RD estimates at conventional levels. For Hispanic prisoners, it appears commissioners were particularly more likely to emphasize the negative aspects of care/harm (column 3) and sanctity/degradation (column 6) compared to Black and white prisoners. Meanwhile, the emphasis on authority appears driven by decisions for Black prisoners specifically (column 7), and commissioners also differentially emphasized loyalty in hearings for Black prisoners compared to white and Hispanic prisoners (column 4).

Overall, Marsy’s Law significantly lengthened denial lengths for all prisoners, while commissioners reached for different moral vocabularies to justify those decisions across racial groups. One possibility is that the different language use reflects different circumstances at the time of the hearing or in a person’s history that the commissioners responded to differentially (for example, if they were more likely to penalize substance use issues and those were differentially distributed across groups). Yet these findings comport with prior work that has argued that commissioners hold different conceptions of criminality and moral worth across racial groups (Greene and Dalke, 2021).

5 Conclusion

What is the relationship between public morality and bureaucratic decisions over punishment? To gain new traction on this question, we make use of rich textual data from transcripts of administrative decisions about whether to release someone from prison and we leverage a voter-mandated change that pushed decision-makers towards harsher punishment. In 2008 California voters passed Marsy’s Law,

which imposed a much more severe set of denial length options for parole commissioners, elevated victims to greater symbolic salience in the proceedings, and included numerous features intended to make hearings more punitive. The law is a classic case of “penal populism” (Pratt, 2007), where appeals to popular moral sentiments of condemnation drive increasingly punitive penal policy. We use Natural Language Processing methods and a Regression Discontinuity design to examine how parole commissioners responded to the Law and its mandates, both in the substantive outcomes they arrived at and the reasons they gave for those outcomes.

The empirical results show that the moral tenor and the severity of punishment moved in opposite directions in response to the Law. We find that the new Law did not lead commissioners to grant release more often, while they increased denial lengths to double the pre-Law mean. However, the evidence we provide from a decomposition analysis suggests that commissioners chose denial lengths of lower ordinal rank from among their new options. This behavioral response tempered the mechanical denial length-increasing effect of the Law, which sought to triple denial lengths. Meanwhile, our analysis of commissioner reasoning highlights that commissioners provided significantly shorter explanations for their decisions after the Law. They mentioned the Law itself in almost a quarter of cases, and when explaining their decisions they used less morally-charged language while emphasizing fairness and authority, while leaning into language related to official documentation. Combining our findings on commissioner decision-making with commissioner reason-giving suggests possible unease with the more punitive mandates of the Law. Ironically, the law appears to have made the hearings *more* bureaucratic – with shorter decisions, more invocation of procedure and official documentation, and less morally strident language – at the same time that they became *more* punitive.

Such findings demonstrate the importance of considering statements of public morality and material outcomes together in administrative decision-making. Our approach builds off the theoretical orientation of Mills (1940), who long ago proposed to treat speech as its own form of action. And we draw from neo-Durkheimian approaches to punishment (Smith, 2008; Garland, 1990) to understand official speech in the context of the criminal-legal system as directing and deflecting public moral sentiment. How it does so is central to understanding the nature of contemporary punishment. Our approach and findings have implications for scholars of morality, frontline bureaucracy, and punishment, and raises novel methodological possibilities for future research.

For students of street-level bureaucracy, we propose bringing public morality into the analytic frame when studying how administrators exercise discretion. Front-line administrators have flexibility not only

in how they apply the rules (Maynard-Moody and Portillo, 2010; Lipsky, 2010; McDonnell, 2025), but also in how they voice and justify their decisions. This symbolic dimension of decision-making – how bureaucrats invoke the public interest and give voice to collective moral sentiment (or don't do so) – constitutes an important basis for the authority of state actors (Bourdieu, 1994, 2014). As such, it deserves greater scholarly attention. The criminal-legal system has many highly visible and often-studied points where administrators make pronouncements on the moral appropriateness of punishment for public audiences, from charging decisions announced in courtrooms and press conferences through to decisions about probation and parole. Other state bureaucracies exercising control over people's lives in adjacent domains – immigration, child welfare, or education, to name a few – also contain ritualistic and ceremonial procedures that claim basis in broader currents of public morality and the public interest.

Our findings suggest that the relationship between articulations of condemnation (and deservingness) are not related to the severity (or generosity) of decisions in a simplistic or mechanical fashion. Our data are limited in important regards for seeing the “backstage” of decision-making processes relative to interview studies and ethnography (e.g. Shiff (2021); Zacka (2017); McDonnell (2020)). Yet our methodological approach allows us to analytically bring into focus important symbolic dynamics of frontline bureaucracy and how actors refract and interact with the moral content of law. Bureaucrats are not simply applying neutral rules to individual cases according to personal moral disposition; they are often times literally giving voice to the currents of public moral sentiment that are encoded in and channeled through the rules they are delegated to apply.

Such an observation provides an invitation for further research into the relationship between decisions and the vocabularies of justification that accompany them. We provide a methodological template for how to do so. Recently, sociologists have been engaged in lively debate about the relationship between what people say and what people do (Jerolmack and Khan, 2014; Lamont and Swidler, 2014; Tavory, 2020; Small and Cook, 2023). These debate have centered methodologically around the kinds of evidence scholars can generate through interviews relative to ethnography. Yet our study shows how emerging text-as-data methods, paired with causal inference frameworks, provide novel opportunities to expand our understanding of the relationship between talk and action in cases where we can clearly observe both.

Our findings also hold implications how we understand the changing practices of punishment. Scholars of mass incarceration often emphasize the rise of managerialism over the last 50 years in the daily administration of the criminal-legal system (Feeley and Simon, 1992; Garland, 2001; Wacquant, 2009). The standard account of the so-called “new penology” argues that frontline bureaucrats shifted focus

away from morally-laden questions of retribution or rehabilitation as the number of people churning through the American criminal-legal system swelled, overwhelming administrators with technocratic demands of population management (Feeley and Simon, 1992). Our findings suggest an alternate, complimentary mechanism pushing towards the bureaucratization of punishment. In our case, we observe commissioners becoming more procedurally-oriented in response to political mandates to impose more severe punishments. While speculative, the findings suggest the possibility that the turn away from an explicit animating moral current in the practices of punishment may be more directly related to the rise of increasingly extreme punishments: as administrators were charged with imposing harsher sentences and more draconian policies, they may have worked to increasingly distance their actions from the moral presuppositions of such policies.

This leads us to conclude with the impacts on those seeking release from prison. It is worth noting that the legal change made a significant difference in narrowing opportunities for people to gain release from prison by doubling the average denial length. It also made the reasons for those penalties more procedural and formalistic, dampening its explicit moral content. On the eve of the punitive turn, Christie (1977) offered a warning that procedural-legal processes can strip disputes, injuries, and modes of redress of their inherent moral content in ways that leave all parties worse off. In an era of penal populism (Pratt, 2007), where the symbolic political authority of punishment has shifted away from expertise towards popular will, the findings suggest that the contemporary administrative structure of punishment is ill-suited for effectively animating and realizing any collective moral sentiment. A similar point has been made by advocates of survivor-centered approaches to restorative justice, who argue that the current American legal system largely fails to meet the emotional needs of those harmed by others (Sered, 2019, pages 20-31). Our findings lend themselves to questions about what sorts of structures could more fully express such complex emotional currents without exacting such an extreme human toll.

Notes

¹Unlike federal court judicial opinions, which have a broader public-facing orientation, the key audience for the parole hearing decision is the prisoner, victims, and future commissioners, as well as courts in the event that a decision is legally challenged.

²Over 80 percent of life-sentenced prisoners are serving a sentence for murder or attempted murder (Weisberg et al., 2011, page 15). Other convictions include for kidnapping and sex offenses.

³Commissioners are tasked with assessing whether a prisoner poses a current danger to public safety (California Penal Code section 3041(b)(1)). The regulations governing the hearing proceedings lay out nine factors demonstrating eligibility and six demonstrating ineligibility that commissioners can consider, though they are not obligated to consider any in particular (California Code of Regulations Title 15, Section 2281).

⁴For information on the Law and vote, see: Marsy's Law Ballotpedia Entry.

⁵Ironically, Marsy's Law increased the public's ability to access the transcripts.

⁶Mechanically, this means we map pre-Law denial lengths of 1 to 3, 2 to 5, 3 to 7, 4 to 10, and 5 to 15.

⁷For the pre-Law period this range is 5 (parole denial length options are 0, 1, 2, 3, 4, and 5 in the pre-Law period). For the post-Law period this range is 15 (denial length options are now 0, 3, 5, 7, 10 and 15).

⁸We obtain a similar null result when we consider the frequency of first-person singular and plural pronouns instead of measuring distance between vector representations.

⁹To get a better sense of the words targeted by these measures, see Table A1 for the ten most-frequent words from the positive and negative poles of each pillar.

¹⁰We list these in Table A1.

References

- An, Jisun, Haewoon Kwak, and Yong-Yeol Ahn. 2018. "SemAxis: A Lightweight Framework to Characterize Domain-Specific Word Semantics Beyond Sentiment." In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, edited by Iryna Gurevych and Yusuke Miyao, pp. 2450–2461, Melbourne, Australia. Association for Computational Linguistics.
- Antoniak, Maria and David Mimno. 2018. "Evaluating the Stability of Embedding-based Word Similarities." *Transactions of the Association for Computational Linguistics* 6:107–119.
- Arango, Tim. 2025. "Heart of the Menendez Case: Who Deserves a Second Chance?" *The New York Times* .
- Ash, Elliott, Daniel L. Chen, and Sergio Galletta. 2022. "Measuring Judicial Sentiment: Methods and Application to US Circuit Courts." *Economica* 89:362–376.
- Ash, Elliott, Daniel L. Chen, and Suresh Naidu. 2018. "Ideas have Consequences: The Impact of Law and Economics on American Justice." *Working Paper* .
- Ash, Elliott and Stephen Hansen. 2023. "Text Algorithms in Economics." *Annual Review of Economics* 15:659–688.
- Auyero, Javier. 2012. *Patients of the State: The Politics of Waiting in Argentina*. Duke University Press.
- Aviram, Hadar. 2020. *Yesterday's Monsters: The Manson Family Cases and the Illusion of Parole*. Oakland, CA: University of California Press.
- Beckett, Katherine. 1999. *Making Crime Pay: Law and Order in Contemporary American Politics*. Oxford University Press.
- Best, Rachel Kahn and Alina Arseniev-Koehler. 2023. "The Stigma of Diseases: Unequal Burden, Uneven Decline." *American Sociological Review* 88:938–969.
- Blinder, Alan S. 1973. "Wage Discrimination: Reduced Form and Structural Estimates." *The Journal of Human Resources* 8:436–455.
- Bourdieu, Pierre. 1994. "Rethinking the State: Genesis and Structure of the Bureaucratic Field." *Sociological Theory* 12:1–18.

- Bourdieu, Pierre. 2014. *On the State: Lectures at the College de France 1989-1992*. Malden, MA: Polity Press.
- Brayne, Sarah and Angèle Christin. 2021. “Technologies of Crime Prediction: The Reception of Algorithms in Policing and Criminal Courts.” *Social Problems* 68:608–624.
- Calonico, Sebastian, Matias D. Cattaneo, and Rocio Titiunik. 2014. “Robust Nonparametric Confidence Intervals for Regression-Discontinuity Designs.” *Econometrica* 82:2295–2326.
- Carlson, Jennifer. 2018. “Legally Armed but Presumed Dangerous: An Intersectional Analysis of Gun Carry Licensing as a Racial/Gender Degradation Ceremony.” *Gender & Society* 32:204–227.
- Christie, Nils. 1977. “Conflicts as Property.” *The British Journal of Criminology* 17:1–15.
- Dalke, Isaac. 2024. “I Come before You a Changed Man: “Insight,” Compliance, and Refurbishing Penal Practice in California.” *Law & Social Inquiry* 49:1138–1168.
- Dobbin, Frank, Daniel Schrage, and Alexandra Kalev. 2015. “Rage against the Iron Cage: The Varied Effects of Bureaucratic Personnel Reforms on Diversity.” *American Sociological Review* 80:1014–1044.
- Durkheim, Emile. 2013. *Durkheim and the Law*. Basingstoke, UK: Palgrave Macmillan, second edition edition.
- Enke, Benjamin. 2020. “Moral Values and Voting.” *Journal of Political Economy* 128:3679–3729.
- Feder, Amir, Katherine A. Keith, Emaad Manzoor, Reid Pryzant, Dhanya Sridhar, Zach Wood-Doughty, Jacob Eisenstein, Justin Grimmer, Roi Reichart, Margaret E. Roberts, Brandon M. Stewart, Victor Veitch, and Diyi Yang. 2022. “Causal Inference in Natural Language Processing: Estimation, Prediction, Interpretation and Beyond.” *Transactions of the Association for Computational Linguistics* 10:1138–1158.
- Feeley, Malcolm and Jonathan Simon. 1992. “The New Penology: Notes on the Emerging Strategy of Corrections and Its Implications.” *Criminology* 30:449–474.
- Ferguson, Jason L. 2021. ““There Is an Eye on Us”: International Imitation, Popular Representation, and the Regulation of Homosexuality in Senegal.” *American Sociological Review* 86:700–727.

- Foreman Jr., James. 2017. *Locking Up Our Own: Crime and Punishment in Black America*. New York, NY: Farrar, Straus and Giroux.
- Frandsen, Brigham R. 2017. "Party bias in union representation elections: Testing for manipulation in the regression discontinuity design when the running variable is discrete." In *Regression discontinuity designs: Theory and applications*, pp. 281–315. Emerald Publishing Limited.
- Frimer, J.A., R. Boghrati, J. Graham, and M. Dehgani. 2019. "Moral Foundations Dictionary for Linguistic Analyses 2.0."
- Garfinkel, Harold. 1956. "Conditions of Successful Degradation Ceremonies." *American Journal of Sociology* 61:420–424.
- Garg, Nikhil, Londa Schiebinger, Dan Jurafsky, and James Zou. 2018. "Word embeddings quantify 100 years of gender and ethnic stereotypes." *Proceedings of the National Academy of Sciences* 115:E3635–E3644.
- Garland, David. 1990. *Punishment and Modern Society: A Study in Social Theory*. Chicago, IL: University of Chicago Press.
- Garland, David. 2001. *The Culture of Control: Crime and Social Order in Contemporary Society*. Chicago, IL: University of Chicago Press.
- Gates, Scott and John O. Brehm. 2010. *Working, Shirking, and Sabotage: Bureaucratic Response to a Democratic Public*. Ann Arbor: University of Michigan Press.
- Gentzkow, Matthew, Bryan Kelly, and Matt Taddy. 2019. "Text as Data." *Journal of Economic Literature* 57:535–74.
- Goldberg, Yoav and Omer Levy. 2014. "word2vec Explained: deriving Mikolov et al.'s negative-sampling word-embedding method." *arXiv:1402.3722 [cs, stat]* .
- Goodman, Philip. 2008. "'It's Just Black, White, or Hispanic': An Observational Study of Racializing Moves in California's Segregated Prison Reception Centers." *Law & Society Review* 42:735–770.
- Graham, Jesse, Jonathan Haidt, Sena Koleva, Matt Motyl, Ravi Iyer, Sean P. Wojcik, and Peter H. Ditto. 2013. "Chapter Two - Moral Foundations Theory: The Pragmatic Validity of Moral Pluralism." In

- Advances in Experimental Social Psychology*, edited by Patricia Devine and Ashby Plant, volume 47, pp. 55–130. Academic Press.
- Greene, Joss and Isaac Dalke. 2021. ““You’re still an angry man”: Parole boards and logics of criminalized masculinity.” *Theoretical Criminology* 25:639–662.
- Gustafson, Kaaryn. 2013. “Degradation Ceremonies and the Criminalization of Low-Income Women Symposium Issue: Critical Race Theory and Empirical Methods.” *UC Irvine Law Review* 3:297–358.
- Hagan, John, Ron Levi, and Ronit Dinovitzer. 2008. “The Symbolic Violence of the Crime-Immigration Nexus: Migrant Mythologies in the Americas.” *Criminology & Public Policy* 7:95–112.
- Haidt, Jonathan. 2012. *The Righteous Mind: Why Good People are Divided by Politics and Religion*. New York, NY: Vintage Books.
- Haney, Craig, Curtis Banks, and Philip Zimbardo. 1973. “Interpersonal Dynamics in a Simulated Prison.” *International Journal of Criminology and Penology* 1:69–97.
- Hempel, Carrie L. 2010. “Lawrence and Shaputis and Their Impact on Parole Decisions in California Probation & Parole Reform.” *Federal Sentencing Reporter* 22:176–180.
- Herbert, Steve. 2022. “Degradation or Redemption? A Parole Board Polices a Moral Boundary.” *Law & Social Inquiry* pp. 1–21.
- Hirschfield, Paul J. and Katarzyna Celinska. 2011. “Beyond Fear: Sociological Perspectives on the Criminalization of School Discipline.” *Sociology Compass* 5:1–12.
- Hollibaugh Jr., Gary E., Matthew R. Miles, and Chad B. Newswander. 2020. “Why Public Employees Rebel: Guerrilla Government in the Public Sector.” *Public Administration Review* 80:64–74.
- Jerolmack, Colin and Shamus Khan. 2014. “Talk Is Cheap: Ethnography and the Attitudinal Fallacy.” *Sociological Methods & Research* .
- Johnson, Eric J., Suzanne B. Shu, Benedict G. C. Dellaert, Craig Fox, Daniel G. Goldstein, Gerald Häubl, Richard P. Larrick, John W. Payne, Ellen Peters, David Schkade, Brian Wansink, and Elke U. Weber. 2012. “Beyond nudges: Tools of a choice architecture.” *Marketing Letters* 23:487–504.
- Kahneman, Daniel and Amos Tversky. 1984. “Choices, values, and frames.” *American Psychologist* 39:341–350.

- Kitagawa, Evelyn M. 1955. "Components of a Difference Between Two Rates." *Journal of the American Statistical Association* 50:1168–1194.
- Kozlowski, Austin C., Matt Taddy, and James A. Evans. 2019. "The Geometry of Culture: Analyzing the Meanings of Class through Word Embeddings." *American Sociological Review* 84:905–948.
- Kutateladze, Besiki L., Nancy R. Andiloro, Brian D. Johnson, and Cassia C. Spohn. 2014. "Cumulative Disadvantage: Examining Racial and Ethnic Disparity in Prosecution and Sentencing." *Criminology* 52:514–551.
- Lamont, Michèle and Ann Swidler. 2014. "Methodological Pluralism and the Possibilities and Limits of Interviewing." *Qualitative Sociology* 37:153–171.
- Lara-Millán, Armando. 2014. "Public Emergency Room Overcrowding in the Era of Mass Imprisonment." *American Sociological Review* 79:866–887.
- Lerman, Amy E. and Vesla M. Weaver. 2014. *Arresting Citizenship: The Democratic Consequences of American Crime Control*. Chicago, IL: University of Chicago Press.
- Lipsky, Michael. 2010. *Street-Level Bureaucracy: Dilemmas of the Individual in Public Service*. New York, NY: Russell Sage Foundation, 30th anniversary edition edition.
- Maynard-Moody, Steven and Michael Musheno. 2012. "Social Equities and Inequities in Practice: Street-Level Workers as Agents and Pragmatists." *Public Administration Review* 72:S16–S23.
- Maynard-Moody, Steven and Shannon Portillo. 2010. "Street-Level Bureaucracy Theory." In *The Oxford Handbook of American Bureaucracy*, edited by Robert F. Durant, p. 0. Oxford University Press.
- McDonnell, Erin Metz. 2020. *Patchwork Leviathan: Pockets of Bureaucratic Effectiveness in Developing States*. Princeton: Princeton University Press.
- McDonnell, Erin Metz. 2025. "Bureaucracy in Action: The Sociology of Public Administration." *Annual Review of Sociology* 51:91–211.
- Mendelsohn, Julia, Yulia Tsvetkov, and Dan Jurafsky. 2020. "A Framework for the Computational Linguistic Analysis of Dehumanization." *Frontiers in Artificial Intelligence* 3.
- Mercier, Hugo and Dan Sperber. 2017. *The Enigma of Reason*. Cambridge, MA: Harvard University Press.

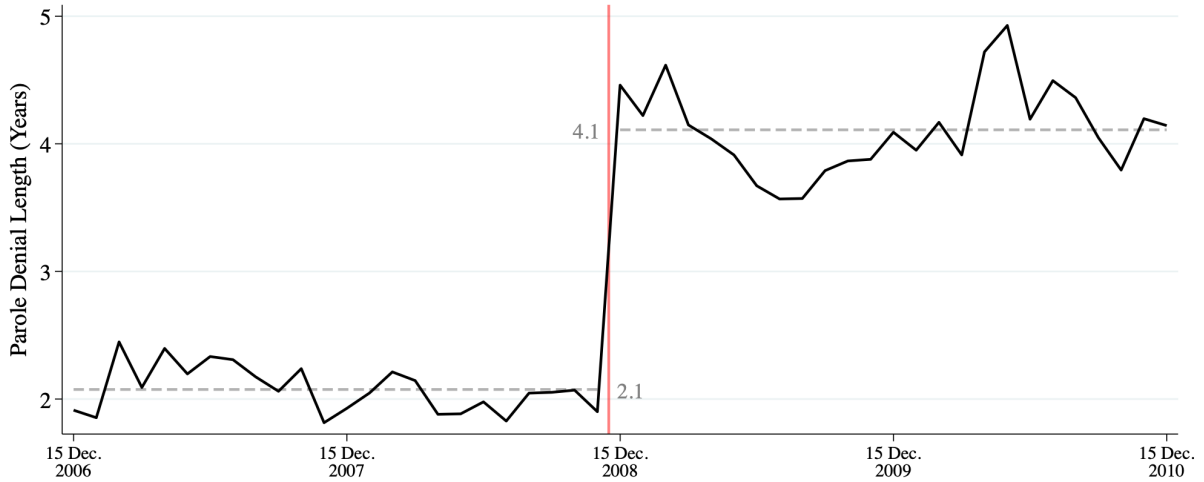
- Mikolov, Tomas, Kai Chen, Greg Corrado, and Jeffrey Dean. 2013. "Efficient Estimation of Word Representations in Vector Space." *arXiv:1301.3781 [cs]* .
- Mills, C. Wright. 1940. "Situated Actions and Vocabularies of Motive." *American Sociological Review* 5:904–913.
- Mohammad, Saif. 2018. "Obtaining Reliable Human Ratings of Valence, Arousal, and Dominance for 20,000 English Words." In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, edited by Iryna Gurevych and Yusuke Miyao, pp. 174–184. Association for Computational Linguistics.
- Muhammad, Khalil Gibran. 2010. *The Condemnation of Blackness: Race, Crime, and the Making of Modern Urban America*. Cambridge, MA: Harvard University Press.
- Musheno, Michael Craig and Steven Williams Maynard-Moody. 2003. *Cops, Teachers, Counselors: Stories from the Front Lines of Public Service*. Ann Arbor: University of Michigan Press.
- Neyman, Jerzy. 1923. "On the application of probability theory to agricultural experiments. Essay on principles." *Ann. Agricultural Sciences* pp. 1–51.
- Oaxaca, Ronald. 1973. "Male-Female Wage Differentials in Urban Labor Markets." *International Economic Review* 14:693–709.
- Portillo, Shannon and Danielle S. Rudes. 2014. "Construction of Justice at the Street Level." *Annual Review of Law and Social Science* 10:321–334.
- Pratt, John. 2007. *Penal Populism*. London: Routledge.
- Richie, Beth E. 2012. *Arrested Justice: Black Women, Violence, and America's Prison Nation*. New York, NY: New York University Press.
- Rios, Victor M., Greg Prieto, and Jonathan M. Ibarra. 2020. "Mano Suave–Mano Dura: Legitimacy Policing and Latino Stop-and-Frisk." *American Sociological Review* 85:58–75.
- Rocklage, Matthew D, Sharlene He, Derek D Rucker, and Loran F Nordgren. 2023. "Beyond sentiment: The value and measurement of consumer certainty in language." *Journal of Marketing Research* 60:870–888.

- Rodman, Emma. 2020. "A Timely Intervention: Tracking the Changing Meanings of Political Concepts with Word Vectors." *Political Analysis* 28:87–111.
- Rodriguez, Pedro L. and Arthur Spirling. 2022. "Word Embeddings: What Works, What Doesn't, and How to Tell the Difference for Applied Research." *The Journal of Politics* 84:101–115.
- Rubin, Donald B. 1974. "Estimating causal effects of treatments in randomized and nonrandomized studies." *Journal of educational Psychology* 66:688.
- Schram, Sanford F., Joe Soss, Richard C. Fording, and Linda Houser. 2009. "Deciding to Discipline: Race, Choice, and Punishment at the Frontlines of Welfare Reform." *American Sociological Review* 74:398–422.
- Sered, Danielle. 2019. *Until We Reckon: Violence, Mass Incarceration, and a Road to Repair*. New York, NY: The New Press.
- Shammas, Victor L. 2019. "The perils of parole hearings: California lifers, performative disadvantage and the ideology of insight." *Political and Legal Anthropology Review* 42.
- Shiff, Talia. 2021. "A Sociology of Discordance: Negotiating Schemas of Deservingness and Codified Law in U.S. Asylum Status Determinations." *American Journal of Sociology* 127:337–375.
- Shiff, Talia. 2024. "Moral Logics of Bureaucratic Indifference." *Annual Review of Law and Social Science* 20:1–15.
- Small, Mario L. and Jenna M. Cook. 2023. "Using Interviews to Understand Why: Challenges and Strategies in the Study of Motivated Action." *Sociological Methods & Research* 52:1591–1631.
- Smith, Philip. 2008. *Punishment and Culture*. Chicago, IL: University of Chicago Press.
- Soss, Joe. 1999. "Lessons of Welfare: Policy Design, Political Learning, and Political Action." *The American Political Science Review* 93:363–380.
- Tavory, Iddo. 2020. "Interviews and Inference: Making Sense of Interview Data in Qualitative Research." *Qualitative Sociology* 43:449–465.
- Thaler, Richard H., Cass R. Sunstein, and John P. Balz. 2013. "Chapter 25. Choice Architecture." In *Chapter 25. Choice Architecture*, pp. 428–439. Princeton University Press.

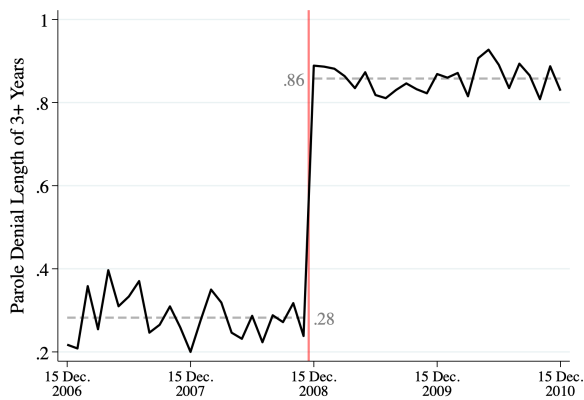
- Tversky, Amos and Daniel Kahneman. 1974. "Judgment under Uncertainty: Heuristics and Biases." *Science* 185:1124–1131.
- Van Cleve, Nicole Gonzalez. 2016. *Crook County: Racism and Injustice in America's Largest Criminal Court*. Stanford, CA: Stanford University Press.
- Wacquant, Loïc. 2009. *Punishing the Poor: The Neoliberal Government of Social Insecurity*. Durham, NC: Duke University Press.
- Watkins-Hayes, Celeste. 2009. *The New Welfare Bureaucrats: Entanglements of Race, Class, and Policy Reform*. Chicago, IL: University of Chicago Press.
- Wattley, Keith. 2013. "Insight into California's Life Sentences." *Federal Sentencing Reporter* 25:1–5.
- Weaver, Vesla M. 2007. "Frontlash: Race and the Development of Punitive Crime Policy." *Studies in American Political Development* 21:230–265.
- Weber, Max. 1946. "Bureaucracy." In *From Max Weber: Essays in Sociology*, edited by H.H. Gerth and C. Wright Mills, pp. 196–244. New York, NY: Oxford University Press.
- Weisberg, Robert, Debbie Mukamal, and Jordan Segall. 2011. "Life in Limbo: An Examination of Parole Release for Prisoners Serving Life Sentences with the Possibility of Parole in California." Technical report, Stanford Criminal Justice Center, Stanford, CA.
- White, Ariel R., Noah L. Nathan, and Julie K. Faller. 2015. "What Do I Need to Vote? Bureaucratic Discretion and Discrimination by Local Election Officials." *American Political Science Review* 109:129–142.
- Young, Kathryne, Debbie Mukamal, and Thomas Favre-Bulle. 2016. "Predicting Parole Grants: An Analysis of Suitability Hearings for California's Lifer Inmates." *Federal Sentencing Reporter* 28:268–277.
- Zacka, Bernardo. 2017. *When the State Meets the Street: Public Service and Moral Agency*. Cambridge, MA: The Belknap Press of Harvard University Press.

Figures

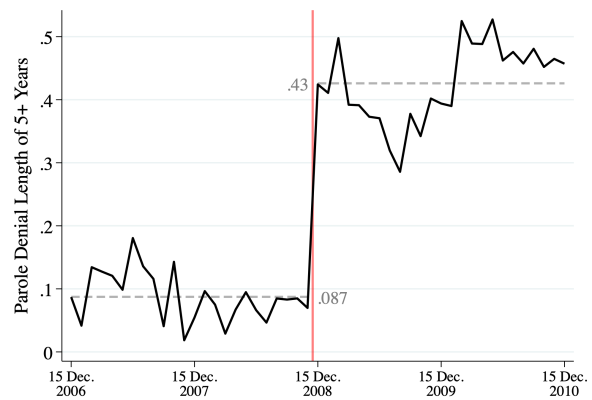
Figure 1: The Impact of Marsy's Law on Parole Denial Length



(a) Parole Denial Length



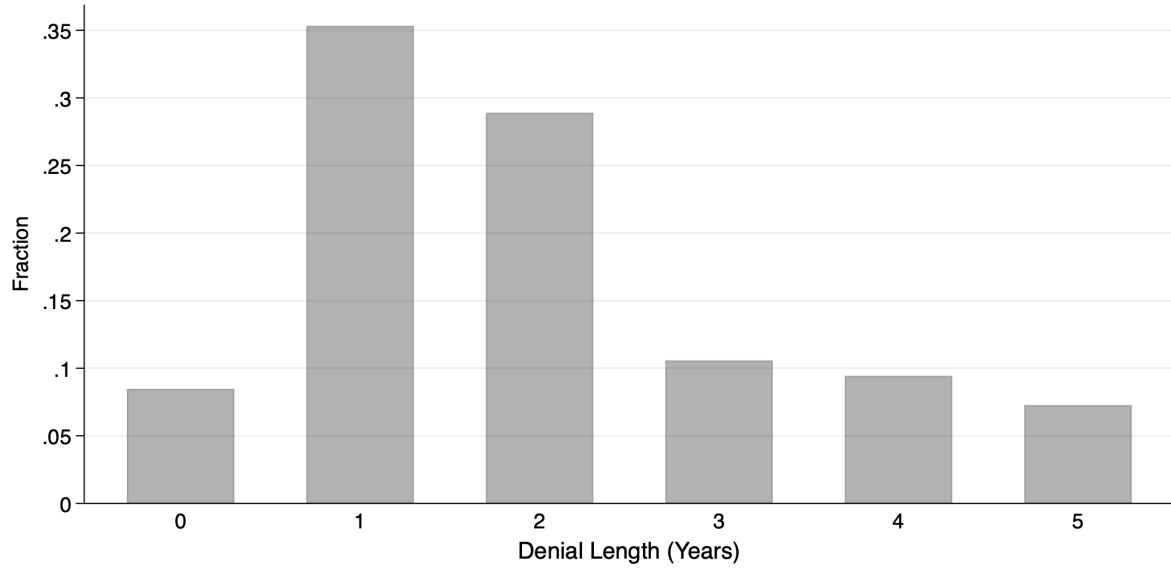
(b) Parole Denial: 3+ Years



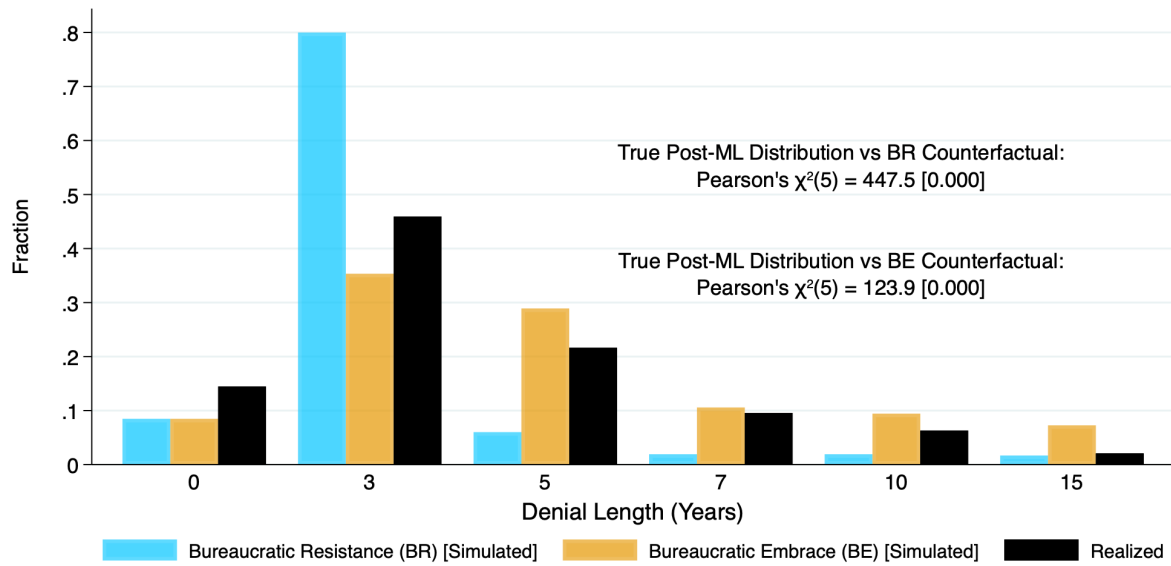
(c) Parole Denial: 5+ Years

Notes: Sample period: January 2007 until December 2010. We collapse the data to the monthly level for these graphs, and start the month on the fifteenth, in order to align our results with the introduction of the Law. In Figure A2, we present the analogous figure for the parole grant rates, for which we observe no discontinuity.

Figure 2: The Realized and Simulated Distributions of Denial Length



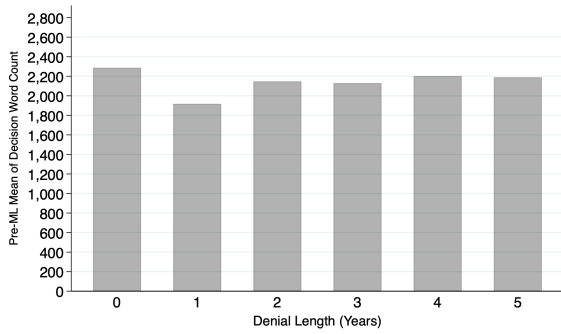
(a) Pre-Marsy's Law Denial Length – Realized



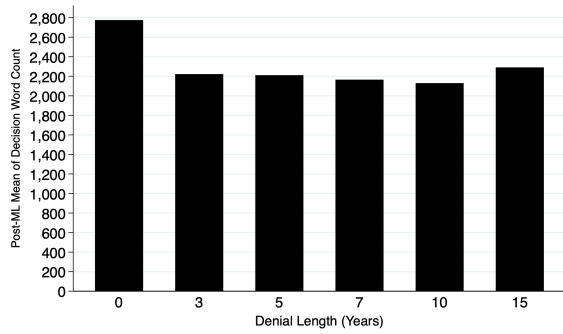
(b) Post-Marsy's Law Denial Length – Realized and Simulated

Notes: The figure is based on parole hearings within a 295 day window around the introduction of Marsy's Law in California. In Figure 2b we present the distribution of denial length under two counterfactual simulations as well as the realized denial length. In the first simulation – the bureaucratic simulation TBC

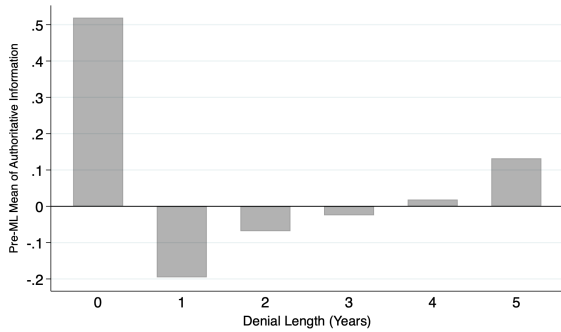
Figure 3: Sample Means of NLP Measures by Period and Denial Length



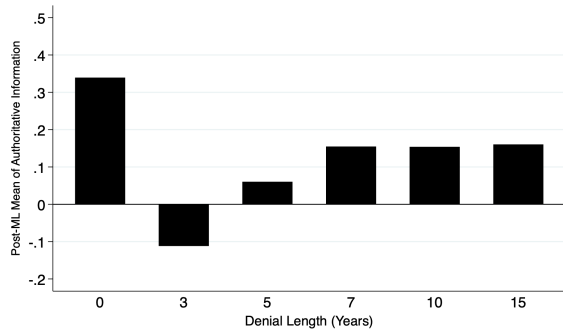
(a) Decision Word Count, Pre-ML



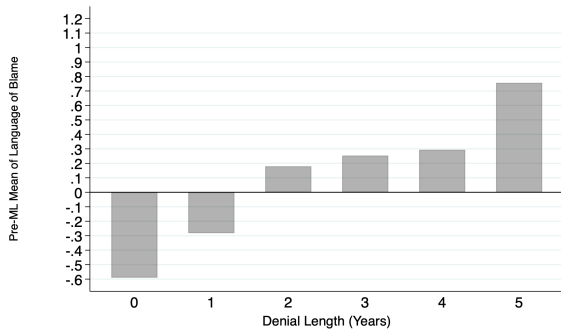
(b) Decision Word Count, Post-ML



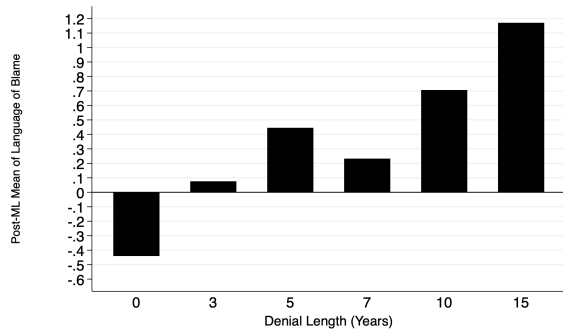
(c) Authoritative Information, Pre-ML



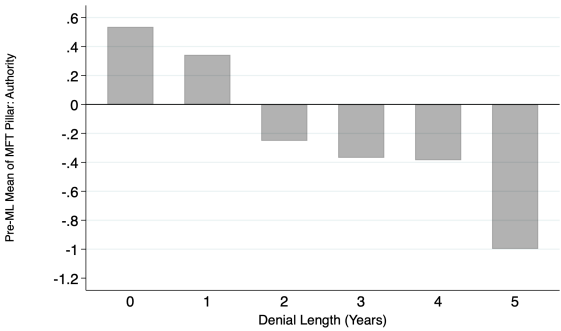
(d) Authoritative Information, Post-ML



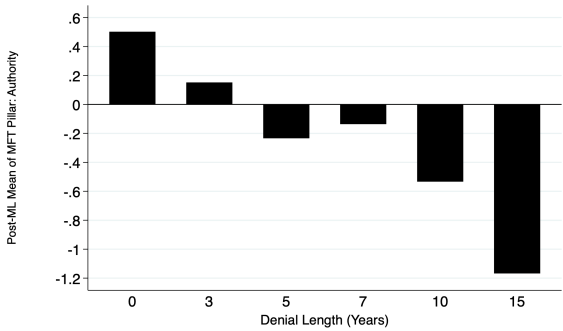
(e) Language of Blame, Pre-ML



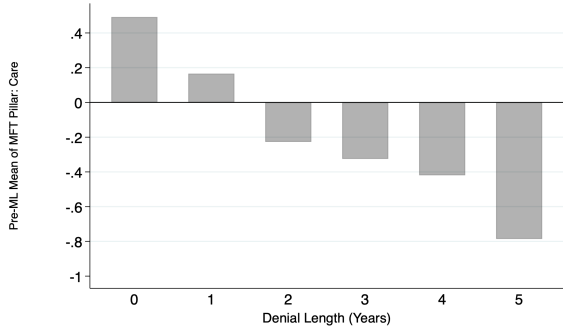
(f) Language of Blame, Post-ML



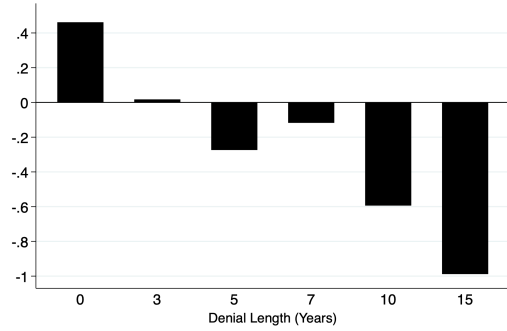
(g) MFT Pillar: Authority, Pre-ML



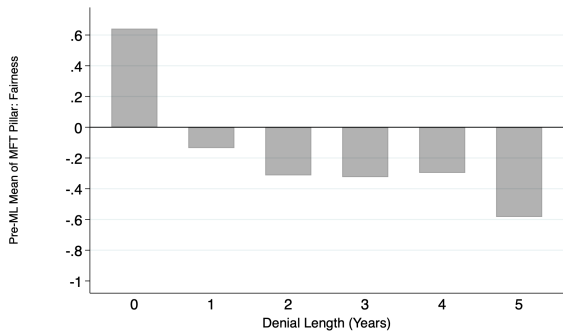
(h) MFT Pillar: Authority, Post-ML



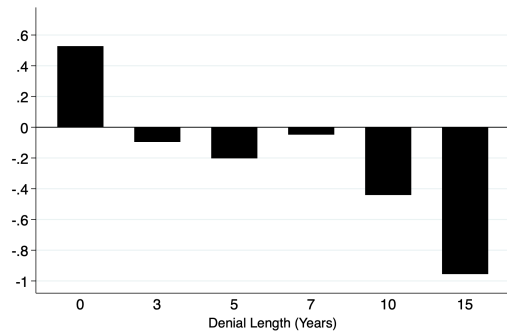
(i) MFT Pillar: Care, Pre-ML



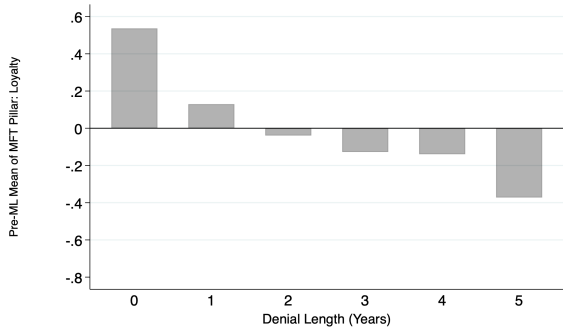
(j) MFT Pillar: Care, Post-ML



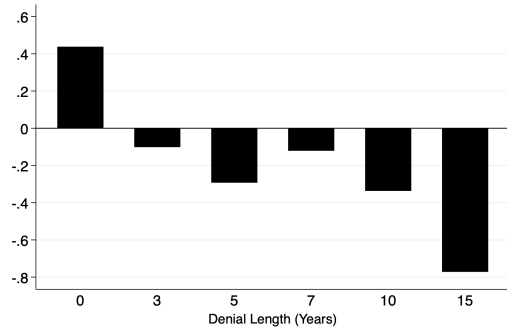
(k) MFT Pillar: Fairness, Pre-ML



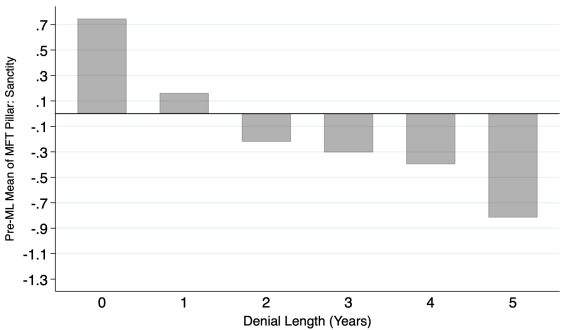
(l) MFT Pillar: Fairness, Post-ML



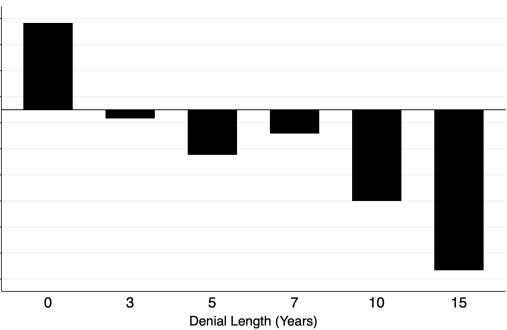
(m) MFT Pillar: Loyalty, Pre-ML



(n) MFT Pillar: Loyalty, Post-ML



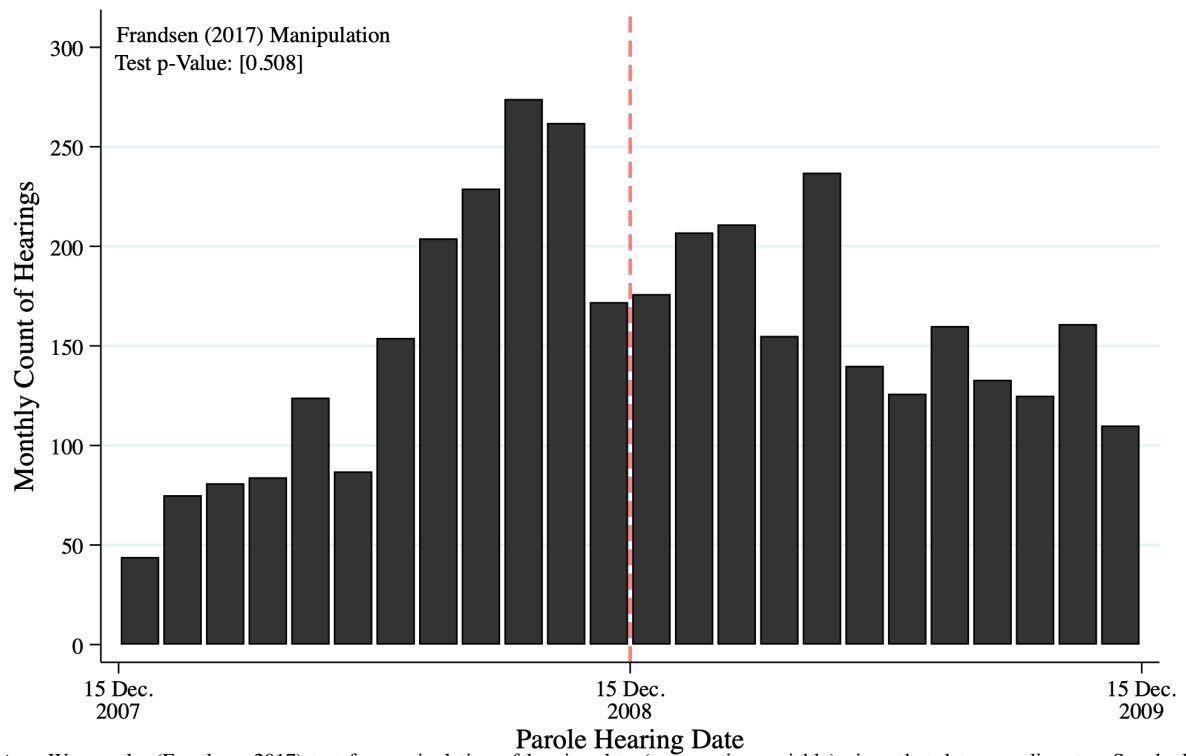
(o) MFT Pillar: Sanctity, Pre-ML



(p) MFT Pillar: Sanctity, Post-ML

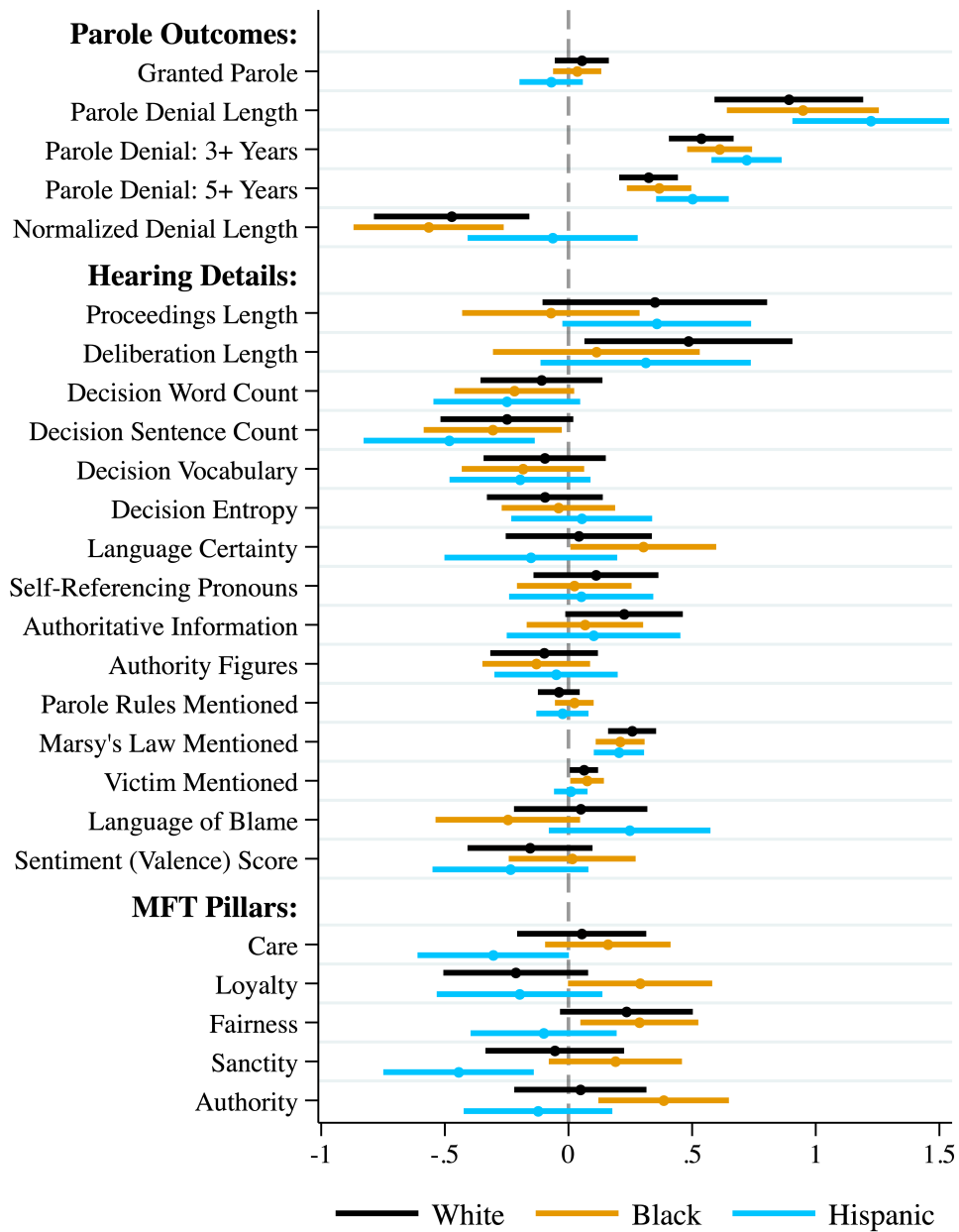
Notes: The figures are based on parole hearings within a 295 day window around the introduction of Marsy's Law in California.

Figure 4: We Document no Evidence of Manipulation of Parole Hearing Date Around the Introduction of Marsy’s Law



Notes: We use the (Frandsen, 2017) test for manipulation of hearing date (our running variable) given that dates are discrete. Standard manipulation tests do not correctly account for this discreteness. In order to implement the test, we need to choose a parameter, $k \geq 0$, which serves as a test leniency parameter – the larger the value of k , the less power the test has to detect manipulation. We choose the strictest possible value: $k = 0$.

Figure 5: Heterogeneity Analysis by Prisoner Race



Notes: The coefficient plot presents coefficients and 95% confidence intervals based on Eicker-White standard errors for the indicator variable Marsy's Law, for both parole outcomes and parole hearings details. All continuous variables are standardized in order to achieve a common scale. Marsy's Law is an indicator variable that takes the value of 1 for hearings from 15 December 2008 and 0 otherwise. Given the temporal structure of commissioner visits to prison, we include prison fixed effects in all specifications. The sample is based on a 293 day window around the introduction of Marsy's Law in California (15 December 2008). We used the approach of Calonico et al. (2014) to calculate optimal bandwidths. In this setting, time is our running variable. All specifications are based on a polynomial of order two in time. The polynomials are included directly, and interacted with the post-Marsy's Law indicator.

Tables

Table 1: Summary Statistics

	(1)	(2)
	Full Sample	Optimal Bandwidth Sample
Sample Size	5,623	3,295
Black	.373	.373
Hispanic	.285	.286
US-born	.843	.844
California-born	.511	.503
Age	63.2 (9.32)	63.5 (9.3)
Time Served (Days)	7,716 (2,357)	7,741 (2,325)
Most Severe Offense: Murder	.769	.768
Most Severe Offense: Sex Crime	.0363	.0349
Granted Parole	.11	.114
Denial Length (Years)	3.17 (2.51)	3.02 (2.51)
Denial Length: 3+ Years	.597	.561
Denial Length: 5+ Years	.266	.233

Notes: Means and standard deviations (in parentheses for continuous covariates) are shown. The Full Sample encompasses the period 01/01/2007-31/12/2010. The Optimal Bandwidth Sample is based on a 295 day window around the introduction of Marsy's Law in California (15 December 2008). We used the approach of Calonico et al. (2014) to calculate optimal bandwidths.

Table 2: Denial Length Decomposition – Pure Bureaucratic Embrace

	(1)	(2)	(3)	(4)
	Denial Length Distribution	Denial Length Choice Set	Counterfactual Simulation?	Denial Length
[1]	Pre-Marsy's Law	Pre-Marsy's Law	No	1.99 (1.35)
[2]	Pre-Marsy's Law	Post-Marsy's Law	Yes	5.28 (3.68)
[3]	Post-Marsy's Law	Post-Marsy's Law	No	4.07 (2.94)

Notes: In Column 4 we present means and standard deviations in parentheses. Marsy's Law is an indicator variable that takes the value of 1 for hearings from 15 December 2008 and 0 otherwise. The sample is based on a 295 day window around the introduction of Marsy's Law in California (15 December 2008). We used the approach of Calonico et al. (2014) to calculate optimal bandwidths. In step 2, we map pre-Marsy's Law denial lengths of 1 to 3, 2 to 5, 3 to 7, 4 to 10, and 5 to 15. This counterfactual simulation assumes a pure bureaucratic embrace of the Law.

Table 3: Denial Length Decomposition Mapping for the Pure Bureaucratic Resistance Response

	(1)	(2)
Original Denial Length	Simulated Denial Length	Probability of Original→Simulated Mapping
0	0	
1	3	
2	3	
3	3	
4	3,5	.5, .5
5	5,7,10,15	.25, .25, .25, .25

Notes: In order to conduct the counterfactual simulation for the pure bureaucratic resistance decomposition, we map denial length under the pre-Law choice set to a new value under the post-Law denial length choice set. This table displays the mapping procedure. We map pre-Marsy's Law denial lengths of 1,2, and 3 to 3. We evenly split pre-Law denial lengths of 4 years between 3 and 5, using a uniform distribution. We evenly split pre-Law denial lengths of 5 years between 5, 7, 10, and 15, using the same uniformly distributed variable.

Table 4: Denial Length Decomposition – Pure Bureaucratic Resistance

	(1)	(2)	(3)	(4)
	Denial Length Distribution	Denial Length Choice Set	Counterfactual Simulation?	Denial Length
[1]	Pre-Marsy's Law	Pre-Marsy's Law	No	1.99 (1.35)
[2]	Pre-Marsy's Law	Post-Marsy's Law	Yes	3.28 (2.14)
[3]	Post-Marsy's Law	Post-Marsy's Law	No	4.07 (2.94)

Notes: In Column 4 we present means and standard deviations in parentheses. Marsy's Law is an indicator variable that takes the value of 1 for hearings from 15 December 2008 and 0 otherwise. Given the temporal structure of commissioner visits to prison, we include prison fixed effects in all specifications. The sample is based on a 295 day window around the introduction of Marsy's Law in California (15 December 2008). We used the approach of Calonico et al. (2014) to calculate optimal bandwidths. In step 2, we map pre-Marsy's Law denial lengths of 1, 2, and 3 to 3. We evenly split pre-Law denial lengths of 4 years between 3 and 5, using a uniform distribution. We evenly split pre-Law denial lengths of 5 years between 5, 7, 10, and 15, using the same uniformly distributed variable. This counterfactual simulation assumes a pure bureaucratic resistance to the Law, and allows for the fact that some of the denial lengths of 5 year in the pre-Law period may have been constrained by the maximum length of 5 years.

Table 5: Testing the RDD Continuity Assumption

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Black	Hispanic	Born in US	Born in CA	Age	Time Served (Days)	Murder	Sex Crime
Marsy's Law	-.0782 (.0477)	.0218 (.0436)	-.0404 (.0351)	-.0806 (.0509)	.385 (.909)	354 (225)	.0271 (.0424)	.0274 (.0197)
\bar{Y}_{PRE}	.379	.282	.848	.504	63.5	7594	.764	.0355
N	3,293	3,293	3,293	3,293	3,293	3,293	3,293	3,293

Notes: *** denotes significance at 1%, ** at 5%, and * at 10%. Eicker-White standard errors in parentheses. Marsy's Law is an indicator variable that takes the value of 1 for hearings from 15 December 2008 and 0 otherwise. Given the temporal structure of commissioner visits to prison, we include prison fixed effects in all specifications. The sample is based on a 295 day window around the introduction of Marsy's Law in California (15 December 2008). We used the approach of Calonico et al. (2014) to calculate optimal bandwidths. In this setting, time is our running variable. All specifications are based on a polynomial of order two in time. The polynomials are included directly, and interacted with the post-Marsy's Law indicator.

Table 6: RDD Evidence on Parole Outcomes

	(1)	(2)	(3)	(4)	(5)
	Granted Parole	Parole Denial Length (Years)	Parole Denial Length of 3+ Years	Parole Denial Length of 5+ Years	Parole Denial Length (Normalized)
Marsy's Law	.00734 (.0313)	2.59*** (.228)	.618*** (.0388)	.396*** (.0379)	-.0912*** (.0227)
\bar{Y}_0	.0847	1.99	.273	.0727	.398
Marsy's Law/ \bar{Y}_0	.0866 (.369)	1.3*** (.115)	2.27*** (.142)	5.44*** (.521)	-.229*** (.057)
Observations	3,293	3,293	3,293	3,293	3,293

Notes: *** denotes significance at 1%, ** at 5%, and * at 10%. Eicker-Huber-White standard errors in parentheses. Marsy's Law is an indicator variable that takes the value of 1 for hearings from 15 December 2008 and 0 otherwise. Given the temporal structure of commissioner visits to prison, we include prison fixed effects in all specifications. The sample is based on a 295 day window around the introduction of Marsy's Law in California (15 December 2008). We used the approach of Calonico et al. (2014) to calculate optimal bandwidths. In this setting, time is our running variable. All specifications are based on a polynomial of order two in time. The polynomials are included directly, and interacted with the post-Marsy's Law indicator.

Table 7: RDD Evidence on Parole Hearing Outcomes

	(1)	(2)	(3)	(4)	(5)
	Proceedings Length (Minutes)	Deliberation Length (Minutes)	Decision Word Count	Decision Sentence Count	Decision Vocabulary Size
Marsy's Law	6.31 (4.21)	5.65** (2.2)	-152** (63.6)	-14.1*** (3.64)	-25.2** (11.5)
\bar{Y}_0	88.9	34	2085	104	606
Observations	2,390	1,842	3,293	3,293	3,293
	(6)	(7)	(8)	(9)	(10)
	Decision Entropy Score	Language Certainty Score	Self-Referencing Pronouns	Authoritative Information & Documentation	Authority Figures
Marsy's Law	-.0092 (.0103)	.0141 (.0174)	.0696 (.073)	.129* (.0726)	-.114* (.0685)
\bar{Y}_0	5.43	6.28	-.0416	-.0361	.141
Observations	3,293	3,293	3,293	3,293	3,293
	(11)	(12)	(13)	(14)	(15)
	Parole Rules Mentioned	Marsy's Law Mentioned	Victim Mentioned	Language of Blame	Sentiment (Valence) Score
Marsy's Law	-.0101 (.0249)	.224*** (.0287)	.0584*** (.0187)	-.0335 (.0838)	-.00205 (.00198)
\bar{Y}_0	.754	0	.938	.0116	.565
Observations	3,293	3,293	3,293	3,293	3,293

Notes: *** denotes significance at 1%, ** at 5%, and * at 10%. Eicker-White standard errors in parentheses. Marsy's Law is an indicator variable that takes the value of 1 for hearings from 15 December 2008 and 0 otherwise. Given the temporal structure of commissioner visits to prison, we include prison fixed effects in all specifications. The sample is based on a 295 day window around the introduction of Marsy's Law in California (15 December 2008). We used the approach of Calonico et al. (2014) to calculate optimal bandwidths. In this setting, time is our running variable. All specifications are based on a polynomial of order two in time. The polynomials are included directly, and interacted with the post-Marsy's Law indicator.

Table 8: RDD Evidence of the Impact of Marsy’s Law on Moral Reasoning

	(1)	(2)	(3)	(4)	(5)
	MFT Pillar 1: Care	MFT Pillar 2: Loyalty	MFT Pillar 3: Fairness	MFT Pillar 4: Sanctity	MFT Pillar 5: Authority
(a) Baseline Measure					
Marsy’s Law	.0246 (.0746)	-.003 (.08)	.173** (.0719)	-.0344 (.0794)	.127* (.0762)
(b) Positive Axis Only					
Marsy’s Law	-.118 (.0809)	-.211*** (.0725)	.0464 (.0817)	-.241*** (.08)	-.0788 (.0704)
(c) Negative Axis Only					
Marsy’s Law	-.0824 (.0733)	-.173** (.0743)	-.121* (.0731)	-.156** (.0723)	-.122* (.0695)
(d) Top 10 Words Only					
Marsy’s Law	-.0127 (.0708)	.104 (.0785)	.173** (.0701)	-.153* (.0796)	.27*** (.0779)
Observations	3,293	3,293	3,293	3,293	3,293

Notes: *** denotes significance at 1%, ** at 5%, and * at 10%. Eicker-White standard errors in parentheses. Marsy’s Law is an indicator variable that takes the value of 1 for hearings from 15 December 2008 and 0 otherwise. Given the temporal structure of commissioner visits to prison, we include prison fixed effects in all specifications. The sample is based on a 295 day window around the introduction of Marsy’s Law in California (15 December 2008). We used the approach of Calonico et al. (2014) to calculate optimal bandwidths. In this setting, time is our running variable. All specifications are based on a polynomial of order two in time. The polynomials are included directly, and interacted with the post-Marsy’s Law indicator.

Table 9: Specific Parole Outcomes, by Race

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Parole Denial Length (Normalized)	Language of Blame	MFT Pillar 1: Care	MFT Pillar 2: Loyalty	MFT Pillar 3: Fairness	MFT Pillar 4: Sanctity	MFT Pillar 5: Authority
White Inmates							
Marsy's Law	-.116*** (.0393)	.0488 (.136)	.0507 (.127)	-.194 (.136)	.221* (.129)	-.0538 (.139)	.0454 (.129)
Black Inmates							
Marsy's Law	-.139*** (.038)	-.243 (.148)	.151 (.123)	.264* (.136)	.27** (.115)	.184 (.133)	.364*** (.128)
Hispanic Inmates							
Marsy's Law	-.0157 (.043)	.245 (.165)	-.289* (.149)	-.18 (.156)	-.0948 (.142)	-.431*** (.151)	-.117 (.145)
p -value: $\alpha^W = \alpha^B$.669	.136	.559	.0144	.77	.205	.0717
p -value: $\alpha^W = \alpha^H$.0767	.345	.0726	.944	.0904	.0576	.39
p -value: $\alpha^B = \alpha^H$.0275	.0235	.0187	.0269	.0398	.00165	.0105

Notes: *** denotes significance at 1%, ** at 5%, and * at 10%. Eicker-White standard errors in parentheses. Marsy's Law is an indicator variable that takes the value of 1 for hearings from 15 December 2008 and 0 otherwise. Given the temporal structure of commissioner visits to prison, we include prison fixed effects in all specifications. The sample is based on a 295 day window around the introduction of Marsy's Law in California (15 December 2008). We used the approach of Calonico et al. (2014) to calculate optimal bandwidths. In this setting, time is our running variable. All specifications are based on a polynomial of order two in time. The polynomials are included directly, and interacted with the post-Marsy's Law indicator.

Appendix

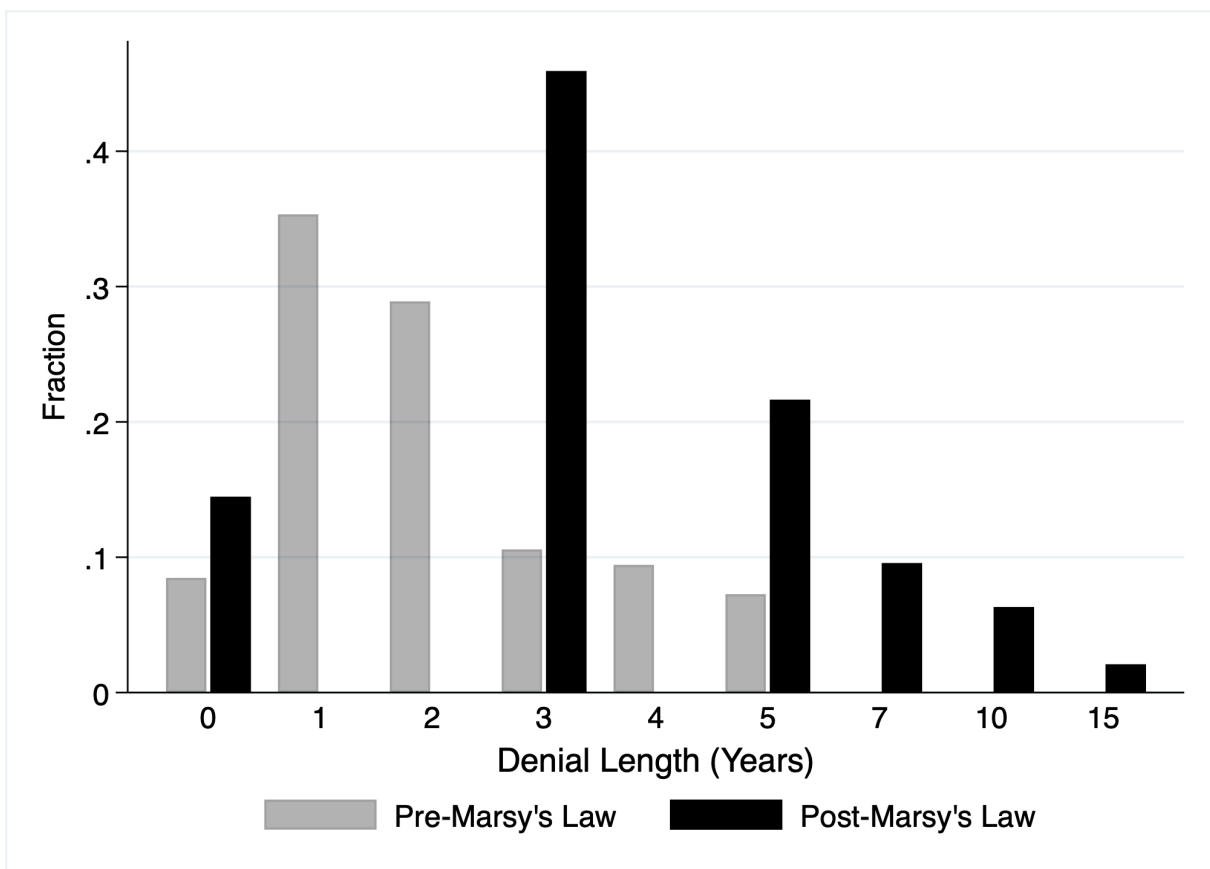
A Additional Results

A.1 Denial Length Distribution and the Law

Figure A1 shows the upward shift in the distribution of denial lengths after Marsy's Law went into effect.

All grants are set to a denial length of 0.

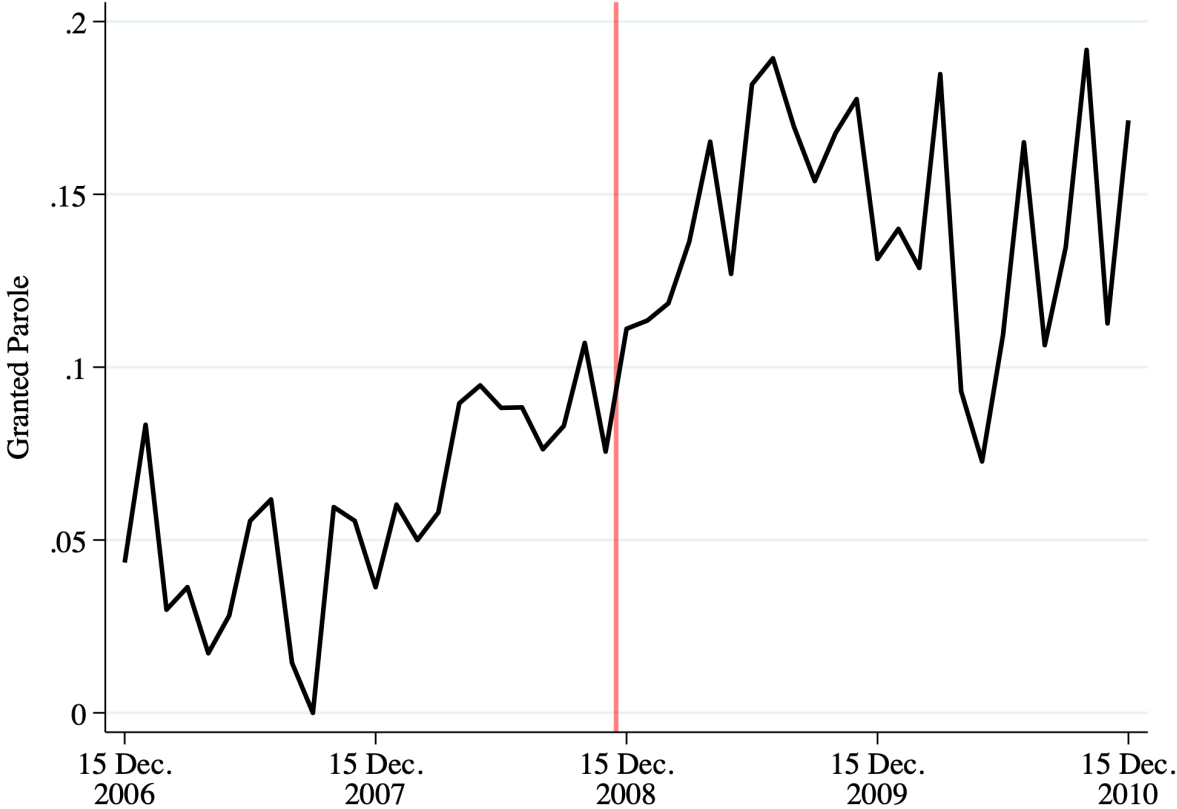
Figure A1: The Distribution of Denial Length and Marsy's Law



A.2 Grant Rate Around Marsy’s Law

Figure A2a shows the portion of hearings per month that resulted in a grant over the study period. It shows a general upward trend in grant rates around the time that the Law went into effect. This secular rising trend in grant rates is important to note, given our central finding that Marsy’s Law specifically had no effect on grants.

Figure A2: The Impact of Marsy’s Law on Parole Grant Rates



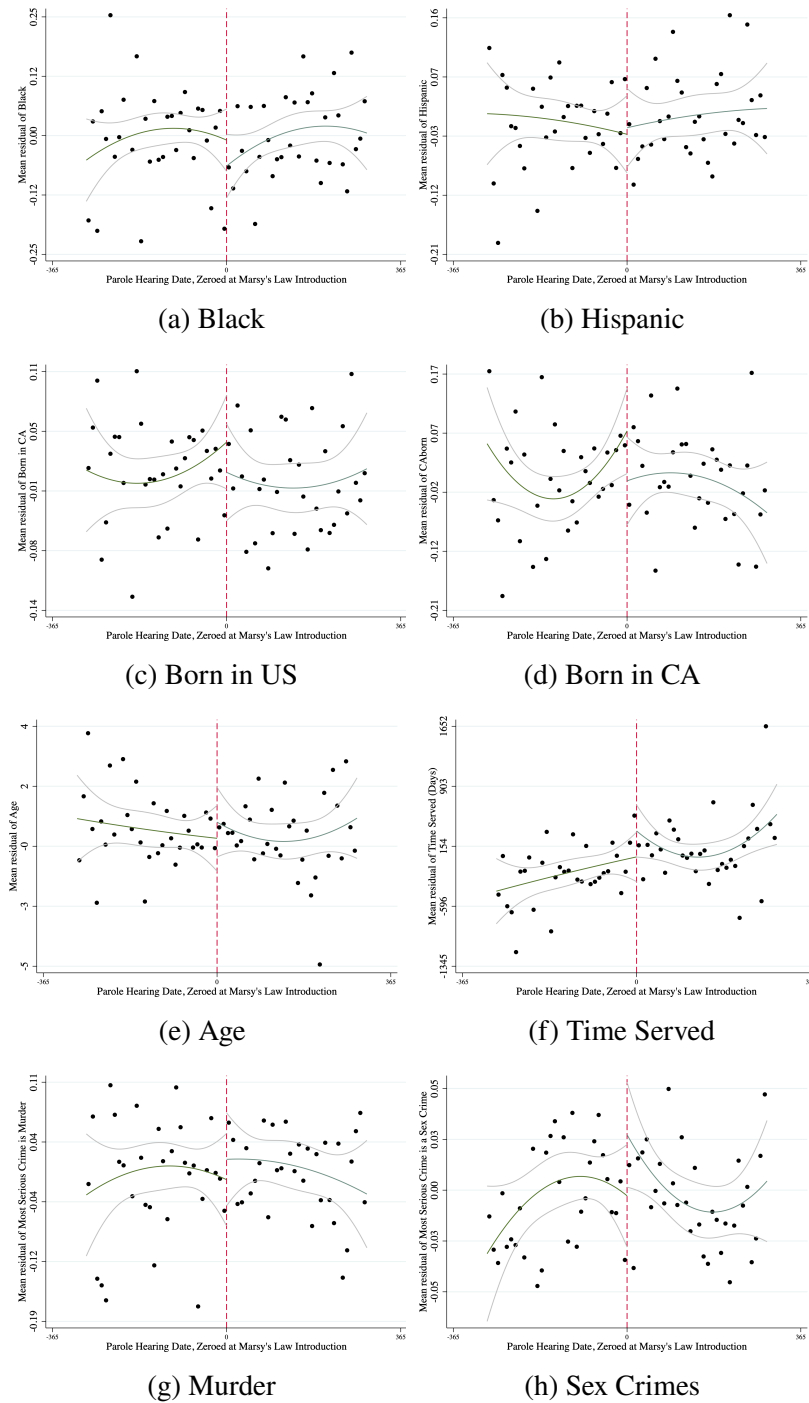
(a) Grant Rate

Notes: Sample period: January 2007 until December 2010. We collapse the data to the monthly level for these graphs, and start the month on the fifteenth, in order to align our results with the introduction of the Law.

A.3 RDD Continuity Assumption: Graphical Evidence

This section plots the distribution of key covariates of people up for parole before and after Marsy’s Law went into effect, with over-time trend lines and confidence intervals. The figures show no detectable shift in the profile of people coming before the board following the implementation of the Law.

Figure A3: The Impact of Marsy’s Law on Key Parole Outcomes



Notes: The sample is based on a 293 day window around the introduction of Marsy’s Law in California (15 December 2008). We used the approach of Calonico et al. (2014) to calculate optimal bandwidths. In this setting, time is our running variable. All specifications are based on a polynomial of order two in time. The polynomials are included directly, and interacted with the post-Marsy’s Law indicator.

A.4 Additional NLP-Based Evidence

In Table A1 we provide the top ten words for each pole, both positive and negative, that constitute the bases of our NLP semantic axes output – the MFT pillar axes and the axis for the language of blame. Recall from Section 2.2 that we manually generated the words for the “blame” axis using dictionaries and thesauruses, while the words included in the MFT pillars come from Frimer et al. (2019).

In Table A2, we present an extended version of the sensitivity analyses in Table 8, splitting the sample by commissioner severity. We return to discuss these findings further in Section A.6.2.

Table A1: Most Frequent Words in NLP-Based Hearing Measures

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
	MFT Pillars											
	Blame		Care		Loyalty		Fairness		Sanctity		Authority	
	Positive	Negative	Positive	Negative	Positive	Negative	Positive	Negative	Positive	Negative	Positive	Negative
1.)	criminal	good	safety	victim	family	enemy	law	theft	clean	drugs	institutional	disorder
2.)	convicted	clear	help	violence	community	infidelity	honest	stolen	church	drug	police	tumultuous
3.)	bad	completed	mother	murder	wife	enemies	laws	lying	religious	sexual	regulations	refused
4.)	sorry	correct	care	victims	group	unfaithful	trust	robbed	bible	alcoholic	respecting	refusing
5.)	wrong	great	child	threat	together	rebel	fair	stole	god	addiction	order	illegal
6.)	guilty	whole	health	suffering	groups	rebellion	justice	stealing	cleaning	defiled	institution	disrespect
7.)	responsible	completed	benefit	cruel	united	outsider	honesty	robbing	faith	alcoholics	arrested	riot
8.)	remorseful	exceptional	vulnerable	killed	fellow	desertion	rights	lied	marriage	horrific	father	unauthorized
9.)	caught	special	helpful	violent	country	deserter	justify	steal	christian	addict	control	refuse
10.)	illegal	honest	helping	kill	familiar	rebels	restitution	segregation	religion	alcoholism	arrest	disrespected

Table A2: Sensitivity Analysis of NLP-Based Hearing Measures, by Commissioner Severity

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
	Lenient Commissioners					Severe Commissioners						
	Language of Blame	MFT Pillar 1: Care	MFT Pillar 2: Loyalty	MFT Pillar 3: Fairness	MFT Pillar 4: Sanctity	MFT Pillar 5: Authority	Language of Blame	MFT Pillar 1: Care	MFT Pillar 2: Loyalty	MFT Pillar 3: Fairness	MFT Pillar 4: Sanctity	MFT Pillar 5: Authority
(a) Baseline Measure												
Marsy's Law	-0.053 (.12)	-.0391 (.108)	.00871 (.115)	-.0372 (.106)	-.111 (.116)	.0162 (.107)	-.0157 (.124)	.0558 (.108)	.0261 (.118)	.389*** (.102)	.0429 (.113)	.22** (.112)
(b) Positive Axis Only												
Marsy's Law	-.103 (.109)	-.0165 (.114)	-.058 (.1)	-.0172 (.116)	-.136 (.119)	.033 (.0938)	-.00462 (.128)	-.23* (.123)	-.328*** (.111)	.135 (.123)	-.351*** (.114)	-.122 (.111)
(c) Negative Axis Only												
Marsy's Law	-.0755 (.128)	.0253 (.103)	-.0552 (.105)	.0219 (.102)	.00309 (.102)	.0101 (.093)	.0148 (.118)	-.167 (.111)	-.288** (.113)	-.254** (.111)	-.32*** (.11)	-.204* (.108)
(d) Top 10 Words Only												
Marsy's Law	-.0259 (.118)	-.0397 (.101)	-.0101 (.114)	-.00038 (.102)	-.0692 (.115)	.0592 (.112)	-.194 (.121)	-.0333 (.103)	.261** (.115)	.33*** (.1)	-.28** (.117)	.489*** (.111)
Observations	1,627	1,627	1,627	1,627	1,627	1,627	1,665	1,665	1,665	1,665	1,665	1,665

Notes: *** denotes significance at 1%, ** at 5%, and * at 10%. Eicker-White standard errors in parentheses. Marsy's Law is an indicator variable that takes the value of 1 for hearings from 15 December 2008 and 0 otherwise. Given the temporal structure of commissioner visits to prison, we include prison fixed effects in all specifications. The sample is based on a 295 day window around the introduction of Marsy's Law in California (15 December 2008). We used the approach of Calonico et al. (2014) to calculate optimal bandwidths. In this setting, time is our running variable. All specifications are based on a polynomial of order two in time. The polynomials are included directly, and interacted with the post-Marsy's Law indicator.

A.5 LDA Topic Modelling Evidence

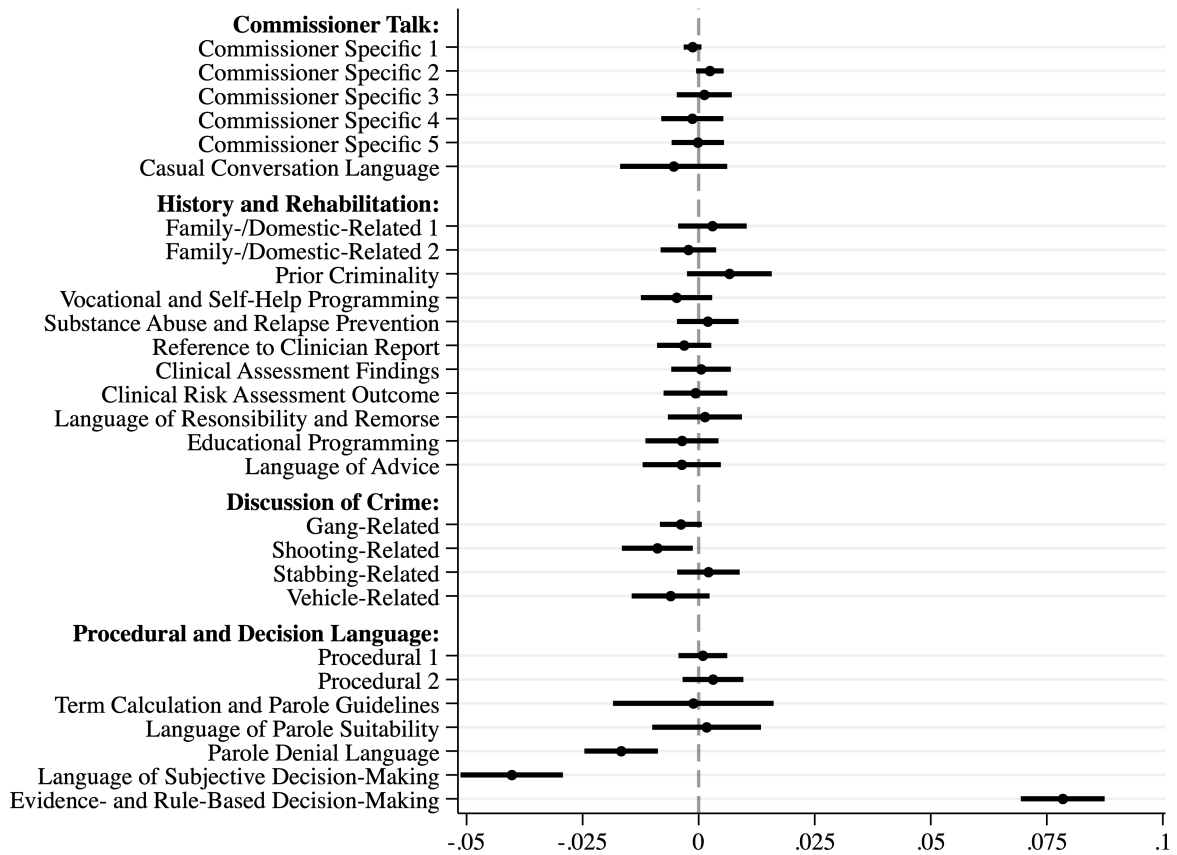
In addition to our semantic axis measures, which take a deductive approach, we implement Latent Dirichlet Allocation (LDA) topic modelling in order to probe commissioner decision language for additional, novel dimensions of speech that may change in response to Marsy’s Law. We train a topic model with Gibbs sampling on all of the transcripts for hearings held between January 1, 2007 and December 31, 2010 using the tomotopy package in Python. We pre-process the text to remove the 100 most common words and any words that appeared in less than 5 percent of the transcripts. Our goal with the topic model is to uncover any changes or trends in the text that may elude our deductive semantic axis approach.

To choose the optimal number of topics we calculate coherence scores for models with topic sizes between 10 and 100 at 10 step increments. This identifies a range of 20 to 30 topics for further investigation. We then manually inspect topic models within that range at 2 step increments to evaluate the interpretability of the topics given our prior knowledge of the hearing contents. Within our selected range, we also tested different values for the alpha parameter (which controls the spread of the distribution of topics over documents), as we expect there to be a high overlap in topics over documents. Through this process we identify a 28-topic model as producing the most interpretable topics across the model. As a final step, both authors independently reviewed and labelled the topics, and then met to harmonize the topic labels in the few cases where there was disagreement.

Our LDA model contains 28 separate topics, which we label and categorize into four key groupings – (i) commissioner talk (verbal tics of commissioners), (ii) history and rehabilitation topics, (iii) discussion of crime, and (iv) procedural and decision language. We present RDD estimates where we consider each topic as an outcome variable in Figure A4. In Table A3 we present the top 10 words for each topic.

We document stability of topics through the Law introduction across the first three categories. It is only in the fourth topic category – on procedural and decision-based language – where we find evidence of responses in commissioner language to the Law. We see declines in the speech-based topics related to the language of denial as well as the language of subjective decision-making. We document a concomitant rise in the frequency of the topic related to evidence- and rule-based decision-making language. Such findings cohere with our previous NLP-based evidence, supporting the interpretation that commissioners appear to responded to the Law by moving away from subjective engagement with the case at hand, and towards a more formalistic, rule-based decision-making language.

Figure A4: LDA Topic Model-Based Evidence



Notes: The coefficient plot presents coefficients and 95% confidence intervals base on Eicker-Huber-White standard errors for the indicator variable Marsy's Law, for both parole outcomes and parole hearings details. All continuous variables are standardized in order to achieve a common scale. Marsy's Law is an indicator variable that takes the value of 1 for hearings from 15 December 2008 and 0 otherwise. Given the temporal structure of commissioner visits to prison, we include prison fixed effects in all specifications. The sample is based on a 293 day window around the introduction of Marsy's Law in California (15 December 2008). We used the approach of Calonico et al. (2014) to calculate optimal bandwidths. In this setting, time is our running variable. All specifications are based on a polynomial of order two in time. The polynomials are included directly, and interacted with the post-Marsy's Law indicator.

Table A3: Most Frequent Words in LDA Topics

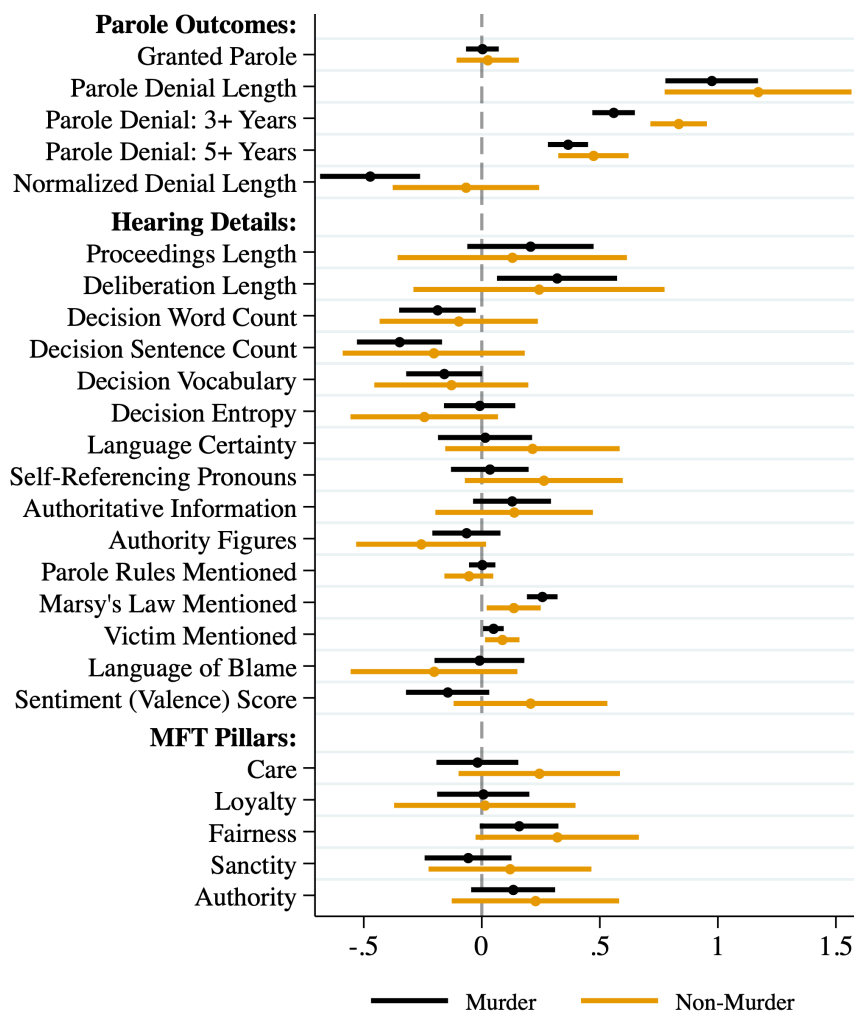
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Comissioner Talk										
Commissioner Specific 1	noting	regards	indicated	obviously	reference	substance	issues	area	particular	disciplinary
Commissioner Specific 2	basically	ended	regarding	stated	able	state	however	terms	anything	different
Commissioner Specific 3	respect	deputy	terms	institutional	thank	remorse	completed	read	programs	unsuitability
Commissioner Specific 4	programming	significant	regards	murder	commended	relates	criminal	since	community	unreasonable
Commissioner Specific 5	indicated	feel	inaudible	however	clearly	able	place	deputy	though	somewhat
Casual Conversation Language	mean	sure	little	somebody	talk	tell	happened	talked	maybe	kind
History and Rehabilitation										
Family-/Domestic-Related 1	wife	child	children	anger	relationship	therapy	issues	women	relationships	daughter
Family-/Domestic-Related 2	although	home	mother	many	kind	feel	house	room	read	letter
Prior Criminality	criminality	robbery	victims	juvenile	failed	possession	criminal	money	jail	theft
Vocational and Self-Help Programming	concern	anger	course	since	management	support	long	participated	completed	vocational
Substance Abuse and Relapse Prevention	substance	plan	relapse	drugs	prevention	drug	able	community	area	programs
Reference to Clinician Report	read	states	page	looked	doctor	sure	section	final	indicate	paroled
Clinical Assessment Findings	noted	considered	respect	issue	doctor	approximately	essentially	encourage	comments	current
Clinical Risk Assessment Outcome	range	moderate	axis	disorder	assessment	personality	overall	score	future	community
Language of Resonsibility and Remorse	responsibility	asked	remorse	statements	version	stated	court	found	evidence	terms
Educational Programming	read	book	give	start	issues	support	five	letters	might	115s
Language of Advice	give	kind	looked	deal	mean	encourage	step	understanding	much	concern
Discussion of Crime										
Gang-Related	gang	member	members	street	activity	gangs	involvement	rival	young	juvenile
Shooting-Related	shot	weapon	victims	shooting	fired	killed	shoot	shotgun	murder	multiple
Stabbing-Related	stabbed	knife	times	death	stab	wounds	human	callous	abused	body
Vehicle-Related	vehicle	police	defendant	told	officer	driving	drove	left	officers	door
Procedural and Decision Language										
Procedural 1	regard	clinician	appear	appears	hours	suitability	significant	criminal	finding	approximately
Procedural 2	terms	individual	regard	individuals	programming	least	provide	another	murder	part
Term Calculation and Parole Guidelines	months	term	base	date	total	governor	review	conditions	degree	murder
Language of Parole Suitability	considerations	months	show	suitability	considered	final	title	grant	provided	fully
Parole Denial Language	following	previous	callous	regard	available	disciplinary	human	separate	participate	failed
Language of Subjective Decision-Making	heavily	suitability	considerations	past	discussed	finding	mental	unsuitability	weighing	state
Evidence- and Rule-Based Decision-Making	require	incarceration	evidence	additional	convincing	weighing	title	clear	code	unsuitable

A.6 Heterogeneity Analysis

A.6.1 Most Serious Offense

In this section we separate hearings for people serving sentences for murder from those serving sentences for other crimes (such as kidnapping or sex offenses). As the majority of cases involve murder, the non-murder cases have much wider confidence intervals. While we observe some slight differences in degree, such as in normalized denial length, or the frequency with which Marsy’s Law is mentioned, all measures appear to move in the same direction for both sets of cases.

Figure A5: Heterogeneity Analysis: Most Serious Offense



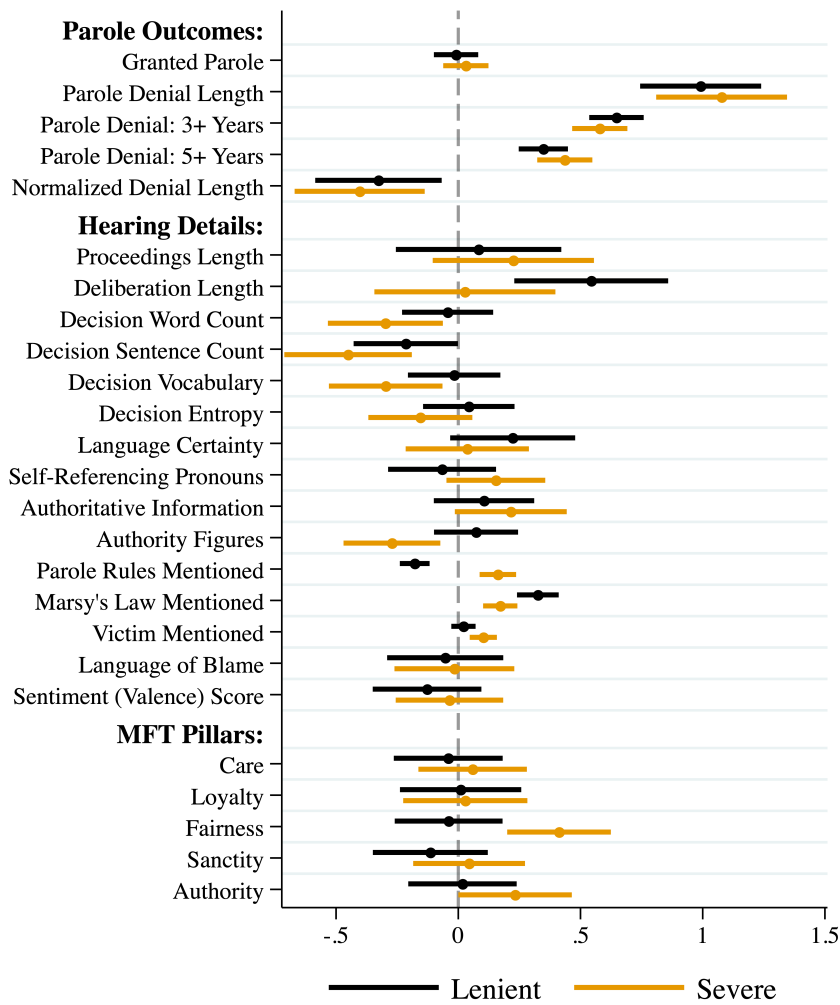
Notes: The coefficient plot presents coefficients and 95% confidence intervals based on Eicker-White standard errors for the indicator variable Marsy’s Law, for both parole outcomes and parole hearing details. All continuous variables are standardized in order to achieve a common scale. Marsy’s Law is an indicator variable that takes the value of 1 for hearings from 15 December 2008 and 0 otherwise. Given the temporal structure of commissioner visits to prison, we include prison fixed effects in all specifications. The sample is based on a 293 day window around the introduction of Marsy’s Law in California (15 December 2008). We used the approach of Calonico et al. (2014) to calculate optimal bandwidths. In this setting, time is our running variable. All specifications are based on a polynomial of order two in time. The polynomials are included directly, and interacted with the post-Marsy’s Law indicator.

A.6.2 Commissioner Severity

In Figure A6, we present sub-sample RD estimates that result from us splitting the commissioners in our sample by ex-ante (i.e., pre-Marsy’s Law) severity. Those with ex-ante denial lengths below the median are labeled “lenient;” their counterparts are labeled “severe.” We highlight four key findings from this analysis. First, commissioners respond homogeneously to the denial length mandates in Marsy’s Law – we find no difference in the RD estimates for any of our outcomes. This set of null findings is surprising given the different pattern of parole denials prior to the implementation of the Law. In the pre-Law period, the mean denial length for severe commissioners was .25 years (or 14%) longer than that of lenient commissioners. Second, it is the severe commissioner group that are the source of our baseline finding of greater reticence post-Marsy’s Law – severe commissioners have statistically significantly lower word counts, lower sentence counts, and use fewer words in their post-Law decisions. The same is not true for the lenient sub-group. Third, the severe commissioners appear to be substituting quantity of words to focus on specific targets – they are statistically significantly more likely to mention parole regulations in general, Marsy’s Law specifically, and to mention victims than they were in the pre-Law period. Fourth, it is severe judges who change their use of moral reasoning in the post-Law period. Our baseline finding of increases in moral reasoning related to the concepts of fairness and authority is driven exclusively by severe commissioners. There is no change in the moral reasoning of lenient commissioners post-Marsy’s Law.

As an extension of this heterogeneity analysis by commissioner severity, in Table A2 we present a variant of our sensitivity analysis of the NLP-based hearings terms (see Table 8) by commissioner severity type. The key conclusion we take from this table is singular and stark – it is severe commissioners who are driving all of the post-Marsy’s Law decline in moral reasoning. We see no changes for lenient commissioners.

Figure A6: Heterogeneity Analysis: Commissioner Severity

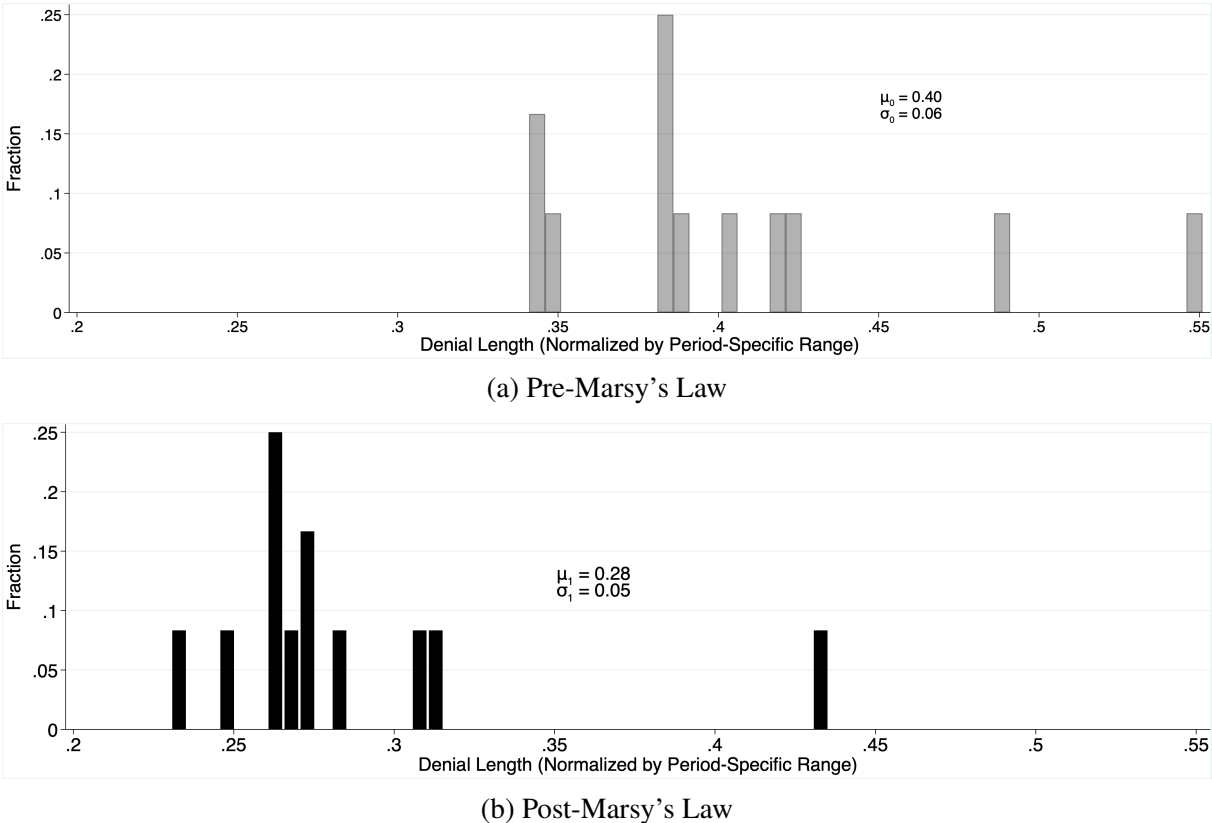


Notes: The coefficient plot presents coefficients and 95% confidence intervals base on Eicker-Huber-White standard errors for the indicator variable Marsy's Law, for both parole outcomes and parole hearings details. All continuous variables are standardized in order to achieve a common scale. Marsy's Law is an indicator variable that takes the value of 1 for hearings from 15 December 2008 and 0 otherwise. Given the temporal structure of commissioner visits to prison, we include prison fixed effects in all specifications. The sample is based on a 293 day window around the introduction of Marsy's Law in California (15 December 2008). We used the approach of Calonico et al. (2014) to calculate optimal bandwidths. In this setting, time is our running variable. All specifications are based on a polynomial of order two in time. The polynomials are included directly, and interacted with the post-Marsy's Law indicator.

A.7 Inter-Commissioner Variation in Denial Length

As a final piece of analysis, we investigate the extent to which inter-commissioner variation in parole denial length changes in response to Marsy’s Law. To make progress on this front, we consider period-specific normalized denial length, in order to have a common yardstick by which to compare commissioner decisions. In Figure A7b we present the mean normalized denial length by commissioner in each period. Recall we have a balanced panel of commissioners, so we are comparing the same commissioners in both periods. These results are the only non-RD estimates we present as part of our core results – here our interest lies in changes across the two periods. Two clear patterns emerge from this analysis. First, as

Figure A7: Inter-Commissioner Variation in Denial Length



Notes: The histograms respectively plot mean (period-specific normalized) denial lengths by parole commissioner for the pre- and post-Marsy’s Law period. As we note above, we restrict our sample to cases heard by commissioners present in both the pre- and post-Law period, hence this is a balanced panel of commissioners. Marsy’s Law is an indicator variable that takes the value of 1 for hearings from 15 December 2008 and 0 otherwise. The sample is based on a 295 day window around the introduction of Marsy’s Law in California (15 December 2008). We used the approach of Calonico et al. (2014) to calculate optimal bandwidths.

we have already documented, commissioners offset the increased length of parole denial length options mandated by Marsy’s Law by handing down lower period-specific denial lengths in the post-Law period. We see the mean normalized denial length drop from .4 to .28 at the commissioner level, in line with our findings in Table 6. Second we see a compression in the distribution of commissioner mean denial

length after the Law goes into effect. There is one outlier commissioner post-Marsys Law, but even with this commissioner, we see a drop in the standard deviation of commissioner parole decisions post-Law. This serves as a complement to the findings from the heterogeneity analysis that we present in Figure 5. While the rank-order of mean commissioner denial length changes slightly across the two periods, not one commissioner in our sample imposes a longer normalized denial length in the post-Law period.