The Reluctant Bureaucrat: Decision-Making and Justification in Shifting Legal Contexts*

Isaac Dalke   Brendon McConnell

September 12, 2024

Abstract

Reason-giving is central to bureaucracies, yet scholars have begun to radically question human capacity for reasoning. Two dominant approaches to cognition take reason-giving as post-hoc justification but diverge on whether the situational context or a person's intuition is more important for understanding the underlying decision-making process. In this paper, we investigate how these competing frameworks bear on bureaucratic reasoning in the criminal-legal system. Specifically, we examine a new law that pushed California parole commissioners towards more punitive decisions about who to release from prison. The situationist framework predicts significantly harsher decisions, while the intuitionist framework suggests commissioners would moderate the law. We use both Natural Language Processing and causal inference methods to analyze 11,704 hearings on either side of the legal change. We provide evidence that the new law unsettled commissioners' intuition, thereby tempering (but by no means negating) its impact. Denial lengths doubled, less than the law intended. The law also included signals to encourage moral condemnation. However, we find that the decisions became less expressive and more bureaucratic: In the post-law period commissioners provided shorter explanations, used less morally-charged language, emphasized case documentation, and invoked the law itself. We consider the implications of bureaucratic reason-giving for legal change.

*Dalke: PhD Candidate, Sociology Department, UC-Berkeley, isaacdalke@berkeley.edu; McConnell: City, University of London, brendon.mcconnell@gmail.com.

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). However tight the rules, bureaucratic actors retain discretion in how to apply them (Lipsky, 2010). A large body of literature in the social sciences makes use of differences in how judges exercise their discretion to study the downstream consequences of bail (Arnold et al., 2018; Dobbie et al., 2018; Arnold et al., 2022) and sentencing (Kling, 2006; Dahl et al., 2014; Harding et al., 2019; Bhuller et al., 2020). This is not just a methodological tool, but of broader political significance: many controversies revolve around what should be the proper basis for bureaucratic action, and whether the stated reasons for action constitute the "real" reasons. In the legal context, examples include what constitutes a legitimate reason for a district attorney to strike a potential juror from a jury (Offit, 2021), or for a police officer to pull over a motorist (Epp et al., 2014), often driven by concerns about racially disparate outcomes. Further complicating this question, such decisions in the criminal-legal context are often not simply procedural, but also explicitly express moral ideas about right and wrong (Durkheim, 2013; Smith, 2008). Central to understanding how decision-makers in the criminal-legal system realize their discretion, then, is a question of how moral reasoning interacts with bureaucratic process. This raises a fundamental question: What is the relationship between moral reasoning, decision outcomes, and the reasons actors give for their decisions in legal bureaucratic processes?

The relationship between moral reasoning and action is a fertile area of inquiry within sociology, with divergent views about how to understand the role of reason in organizing social life. Some have argued that action is best explained by the circumstances of the situation rather than the characteristics of the actor (Jerolmack and Khan, 2014; Mills, 1940; Ross and Nisbett, 1991). We call this the *situationist* approach. From this vantage point, reason-giving is most fruitfully viewed as post-hoc rationalization: what counts as a good reason will be dependent on the specific situation, and so reason-giving is best explained by the situation rather than by characteristics of the individual. As some have noted (Bruch and Feinberg, 2017), this argument is complementary to a large body of empirical work in behavioral economics and psychology that finds that aspects of the "choice architecture" such as the available options or the default option can have a large impact on the choice that individuals make (Thaler, 2016; Kahneman and Tversky, 1984; Tversky and Kahneman, 1974).

In contrast to the situationist approach, what we call the *intuitionist* approach argues to retain a space for moral reasoning as motivating action (Vaisey, 2009). From this view, people hold relatively stable conceptions of what is good and desirable, even if those conceptions may be intuitive, pre-reflexive, or habitual rather than fully articulated (Bourdieu, 1990; Lizardo and Strand, 2010; Martin, 2011). This builds on work in social psychology on moral reasoning that finds that individuals have relatively stable intuitions about what is good and bad, and then reach for different moral reasons to justify those intuitions (Haidt, 2012; Mercier and Sperber, 2017).

The two approaches are not strictly mutually exclusive, as both view reason-giving as a means of justification. But they do point to different drivers of administrative action: organizational context versus personal intuition. The situationist framework heavily emphasizes the setting and decision architecture in shaping the decisions people make, such as anchoring and priming. Meanwhile, the intuitionist framework emphasizes the disposition of the individual. We do not have a clear picture of the relationship between personal intuition and organizational context in practice.

To help fill this lacuna, we investigate the relationship between reason-giving and outcomes in high-stakes bureaucratic settings: decisions about whether to release people from prison. 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 a 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. In 2008, California voters passed a proposition that significantly increased parole denial lengths, set the default option at the maximum denial length, and symbolically re-affirmed the importance of the victim in the proceedings. However, the new law did not change the reasons that commissioners could or should rely on when determining whether to let someone out of prison. How did the harsher substantive outcomes change the reasons commissioners gave for their decisions? An emphasis on context would lead us to expect much harsher moral reasoning: the symbolic framing of victims and more severe choice should push commissioners to be much more condemnatory in their decisions. Yet an emphasis on intuition would lead us to expect commissioners to be much more uncertain in their reason-giving, as they would have less strong intuitions about the "right" outcome under the new conditions.

To empirically investigate these possibilities, we analyze 11,704 parole hearing transcripts for hearings held between January 1, 2007 and December 31, 2010. The transcripts provide a unique opportunity to examine both decision outcomes and the reasons given for those decisions.[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 aspects of the hearing. 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 new law with reticence and caution. The law had no effect on the number of people granted parole, and doubled the average denial length, as its authors intended it to do. Yet commissioners gave shorter denial lengths than if they had simply applied their old relative preferences to the new denial lengths. 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 emphasized fairness and authority more after the new law went into effect, 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 appears to have challenged the intuition of the commissioners about what the right outcome should be, and they adapted by becoming more guarded and reserved in their moral reason-giving — the opposite of the signals the law sent to be harsher and more condemnatory. Ironically, the impact of a law designed to make hearings more morally expressive resulted in making decisions appear more bureaucratic. We take this as evidence of the intuitionist approach, whereby reason-giving is a means of justifying one's intuition about a fair outcome. It is worth noting that we do not believe context is irrelevant — particularly for those who bear the brunt of decisions. What was most determinative for outcomes was, of course, the law itself: commissioners could have been much more severe than they were in practice, but still in the absence of the new law people up for parole would have received significantly more opportunities to gain release. Our evidence suggests that the disposition of the bureaucratic actors responsible for implementing the law muted its effects, but by no means negated it.

The paper also makes broader theoretical and methodological contributions. We build on calls to integrate the interdisciplinary literature on judgment and decision-making into sociology (Bruch and Feinberg, 2017; Vaisey and Valentino, 2018), particularly for understanding legal change and administrative discretion. Our paper suggests the importance of decision-makers' built-up practical intuition for mediating legal and administrative reforms. Bureaucratic routines can build up strong instincts for navigating decision-making, and change to bureaucratic practice thus have to contend with the prior experience and disposition of the bureaucratic actors responsible for implementing it. While we examine a legal reform that was punitive in nature, we expect that administrators would be similarly hesitant in implementing reforms towards greater leniency. This is a fertile area for future investigation. Methodologically, our work also 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).

The paper is structured as follows. We begin by considering the situationist and intuitionist accounts of decision-making, how they may operate in a structured bureaucratic environment, and how they link decision-making to reason-giving. We then present Marsy's Law and how the two approaches predict administrative decision-makers would respond to the changes imposed by the law. After that, we move into the data, research design, and empirical findings, before concluding with implications for bureaucratic decision-making and legal change.

2 Literature Review

Social life is shot through with reasons, accounts, and explanations of past action. This is nowhere more so than in bureaucracies, which are supposed to be governed by a consistent and transparent logic, the apex of reason (Weber, 1946). The ideal of reasoning within bureaucracies meshes with classical notions of reasoning as the logical, objective application of principles to concrete cases in a consistent manner that can be clearly articulated. Yet over the last 50 years, an overwhelming body of evidence has grown that individuals are not adept, as a whole, at reasoning (Mercier and Sperber, 2017). Meanwhile, a long tradition in sociology has questioned the relationship between a person's actions and the reasons a person gives for their actions (Jerolmack and Khan, 2014; Mills, 1940). Yet this has led to different views on how reasons are wielded and how they connect back to underlying cognitive processes. In what follows, we contrast two approaches: one that emphasizes the specific situation in eliciting reasons, and another that emphasizes personal intuition. We consider how both general accounts of reason-giving may operate within a bureaucratic organization. While the two are not mutually exclusive, we consider how they point in opposite directions for predicting how bureaucratic actors may respond to externally-imposed change.

2.1 The Significance of the Situation

First, some have drawn attention to how context shapes decision-making, what we call the 'situational' approach. The core argument is that the social environment is more explanatory of individual decision-making than the intrinsic motivation of individuals (Ross and Nisbett, 1991). An emphasis on the decision-making context or situation in shaping outcomes is now strongly associated with behavioral economics (Thaler, 2016). Yet sociologists have also taken a renewed interest in delineating the role of the situation in shaping action. Bruch and Feinberg (2017) write, "By leveraging insights on how contextual factors and aspects of choice problems influence decision strategies, sociologists can better pinpoint how, why, and when features of the social environment trigger and shape human behavior" (209; see also Hitlin and Vaisey (2013)).

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). Both framing and anchoring involve the presentation of the specific task, which interacts with broader social norms and cultural meanings. For example, survey designers have long grappled with social desirability bias, where actors may choose answers on surveys that they believe others would find acceptable, even if false (Phillips and Clancy, 1972). Social desirability bias is not only about the specific options an actor has to choose from, but also the broader social context in which they are making a decision.

Some of this work focuses specifically on repetitive and consequential bureaucratic decision-making. One of the most famous (and contested) studies of situational effects comes from parole board decisions in Israel: Danziger et al. (2011) found that decision-makers were least likely to grant parole right before they took a scheduled break, and most likely immediately after. Focusing on the U.S. criminal-legal system, Pryor et al. (2020) argue that racially discriminatory policing practices are better explained by the contexts of frontline police work that induce discriminatory behavior rather than the disposition of individual police officers. Meanwhile, Rachlinski and Wistrich (2018) find strong anchoring effects in how judges reason about hypothetical scenarios, while Englich (2006) summarizes numerous studies that found evidence of prosecutor's recommendations anchoring judges' ultimate decisions in sentencing.

Much of this work focuses only on the outcomes of decision-making, not the reasons given for a particular decision. For example, the study of parole decisions by Danziger et al. (2011) did not consider whether the reasons commissioners gave changed by time of day, only outcomes. Meanwhile, some have found that simply the requirement to give a reason can change the decisions individuals make (Hsee, 1999; Wilson et al., 1993). Yet the situationist framework predicts that reason-giving is a product of context as well, if we think of reason-giving as a distinctive decision-making process in itself (Jerolmack and Khan, 2014). From this perspective, the vocabulary of reason-giving belongs to the situation rather than the individual. Mills (1940, p. 909-910) went so far as to argue that researchers should treat all post-hoc accounts of action as justifications matched to a situation, rather than articulations of an inner state.

In sum, the situationist approach expects the decisions actors make, and the reasons they offer, to be strongly influenced by both the design of the task and the broader social context. If the situation changes, then the decision and explanation offered for that decision should change accordingly. In contrast, the intuitionist approach turns to the stable proclivities of the individual to explain decision-making and reason-giving. We consider this approach next.

2.2 The Persistence of Intuition

The intuitionist approach provides an alternate perspective on decision-making and its relationship to reason-giving. The core argument is that (moral) reasoning is largely based on the snap response that an individual has to a situation — is it good or bad, appealing or repulsive — and reasons come when individuals seek to articulate their snap response. In this view, reasoning is still justificatory, but connected back to individual disposition rather than simply social context. This pre-reflexive, instinctual account of decision-making and reason-giving is well-aligned with sociological accounts of habitual and instinctual action that emerge out of an actor's prior experience of the world (Bourdieu, 1990; Lizardo and Strand, 2010; Martin, 2011, p. 262-265). Some of the work on moral intuition points to the ways that external environments can shape reason-giving (Schnall et al., 2008, e.g.). But individuals' intuitions can prove quite stable over time (Vaisey, 2009), and so understanding outcomes points back to the individual as much as to the situation. Meanwhile, reason-giving is a way of articulating, evaluating, and attempting to persuade others of those intuitions (Mercier and Sperber, 2017).

The social psychologist Jonathan Haidt (2012) has put forward perhaps the most influential framework for linking moral reasoning and intuition across the social sciences. Haidt and co-authors originally identified five distinct 'pillars' of moral reasoning: sanctity/degradation, loyalty/betrayal, fairness/cheating, authority/subversion, and care/harm (Graham et al., 2013). The cultural prevalence of different pillars varies across place and time, and vary at the individual level by a person's disposition and upbringing. Each pillar has a distinct vocabulary (Frimer et al., 2019). The Moral Foundations Dictionary has found wide application in NLP-based studies studying moral reasoning, with scholars deploying them to measure disease stigma (Best and Arseniev-Koehler, 2023), dehumanization (Mendelsohn et al., 2020), and moral appeals in U.S. congressional speeches (Enke, 2020), among other topics.

To our knowledge, the moral foundations framework has not been rigorously applied to reasoning in bureaucratic criminal-legal settings. American law is highly rationalized, formally governed by the application of strict rules and the reasoned weighing of evidence with logic. However, dating back to Durkheim (2013, chap. 3), sociologists have long noted that punishment contains a deep moral dimension (Smith, 2008). Today, scholars point specifically to contemporary parole boards as sites of moral expression and boundary-making (Aviram, 2020; Greene and Dalke, 2021; Herbert, 2022; Shammas, 2018). Thus, we would expect decisions about who to punish, and for how long, to involve similar processes of moral evaluation and justification to what scholars have found in other social domains. The Moral Foundations Theory framework provides us with a starting point for empirically investigating the moral reasoning in parole decisions with a theoretical account of the underlying decision-making process (Hitlin and Vaisey, 2013, p. 57).

In contrast to the situationist approach, the intuitionist account re-inserts the individual into the decision-making process. How do the two approaches expect bureaucratic decision-makers to adapt to externally-imposed changes? We will explore this question in a following section. First, we introduce our case.

2.3 The Case: Marsy's Law and the Parole Board

In this paper, we investigate the link between reason-giving and outcomes for parole board commissioners making decisions about who to release from prison in California. In 2008, a voter-passed initiative would dramatically alter the stakes of commissioners' decisions, while leaving untouched the formal reasons they could rely on to reach a decision. We want to know how the commissioners responded to this externally-imposed change to the legal and organizational context. In this section, we provide important background on the parole board and describe salient features of the 2008 law.

Parole decisions are high stakes. California's sentencing system is largely determinate, with only a select number of sentences eligible for release through the state's parole board. While parole eligibility has begun to expand since 2016, during our study period only people serving an indeterminate life sentence were eligible for a hearing. Such sentences set a minimum date for parole eligibility, but have no fixed release date (for example, 15-to-life). 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). Over 80 percent are serving a sentence for murder or attempted murder (Weisberg et al., 2011, page 15). 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). For this group, the only way out of prison goes through the parole board.

The hearings are also highly bureaucratic. The state's Board of Parole Hearings (BPH) oversees hearings, which typically take place in the prison where the parole candidate is incarcerated. The hearing is led by a Presiding Commissioner who is a political appointee, and a Deputy Commissioner who is an administrative employee. The two engage in extensive questioning of the prisoner, going over a person's life before prison, commitment offense, time in prison, and plans if released. The prisoner is entitled to legal representation to guard their rights in the proceedings. At the end of the proceedings, a district attorney from the sentencing county may also make a statement, as well as victims, victims' next of kin, and/or victim representatives. The proceedings can take hours. The commissioners then enter into a private off-the-record deliberation period where they arrive at a decision. After arriving at a decision, they enter back onto the record to announce their decision and the reasoning behind it to all present. If commissioners deny parole, they also set the length of time before a person is eligible for another hearing (known as the denial length).

While the hearings follow a consistent bureaucratic form, the decisions themselves are highly discretionary. Commissioners are supposed to assess 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 (California Code of Regulations Title 15, Section 2281). However, the regulations give commissioners wide latitude in applying the factors: Commissioners don't have to consider any factors in particular, and only have to meet a standard of 'some evidence' to support their decision (Wattley, 2013, page 2).

Against this backdrop, on November 4, 2008 California voters passed Proposition 9 (also known as the California Victims' Bill of Rights Act of 2008 or "Marsy's Law"). The proposition fits into the state's long tradition of passing more punitive criminal laws in the name of "victim's rights" (Young, 2016, pages 446-448, 470). The ballot initiative was spearheaded by Henry Nicholas, a tech entrepreneur whose sister Marsalee (or "Marsy") had been murdered and whose assailant subsequently received a parole-eligible life sentence (Aviram, 2020, pages 52-54). The proposition wrote a Victims' Bill of Rights into the California state constitution and included two major provisions impacting parole hearings.

First, the new law re-framed hearings to emphasize the consequences of decisions for both individual victims and "victims" in general. As part of a new Victim's Bill of Rights, the law increased the ability for victims, victims' family members, and designated victim representatives to attend the hearing and make a statement as part of the proceedings. This change in the framing would place greater emphasis on specific victims and families in specific hearings. Meanwhile the passing of a victims' bill of rights signaled the symbolic importance of "the victim" more generally.

Second, the law changed the choice architecture of decisions. More specifically, it made both the minimum and maximum denial lengths three times more severe than before the law went into effect, and moved the default denial length. Prior to Marsy's Law, commissioners could deny a prisoner for 1, 2, 3, 4, or 5 years, and were supposed to begin with a default 1 year denial before moving up to longer denial lengths. After Marsy's Law, commissioners could deny a prisoner for 3, 5, 7, 10, or 15 years, and were supposed to begin with a default of 15 years and work down. This new choice set and default choice were designed to push the commissioners to issue much longer denial lengths.

The parole board began implementing the new law on December 15, 2008.[2] In interviews with commissioners two years after the law was implemented, Young (2016, 462-466) found that commissioners supported the longer maximum denial length of 15 years, but were concerned that the absence of shorter minimum denial lengths was unfair to people who were close to parole. This suggests that commissioners were ambivalent about the legal changes brought on by the law. However, we do not know how this impacted the hearings themselves, and this sort of retrospective interview data is the exact type of evidence that the situationist approach distrusts (Jerolmack and Khan, 2014). How did the new legal and organizational context change the outcomes commissioners arrived at, and the reasons they gave for those outcomes?

2.4 Empirical Expectations

The situationist and the intuitionist approaches provide diverging expectations for how parole board commissioners would have responded to this shift. The two approaches are clearly not mutually exclusive: individuals can both be influenced by their situation and by their intuition. Yet whereas the situationist approach suggests that changes to the context will dictate outcomes, the intuitionist approach instead emphasizes the persistence of personal intuition across changes to context. This leads to opposite expectations about how parole commissioners would have responded to Marsy's Law.

The law mandated much more punitive outcomes, and sent broader social signals condemning prisoners while highlighting the sanctity of victims. It also implemented a series of changes to the specific choice architecture – shifting the choice set and default – that a behavioral economics approach predicts would lead to much more severe decisions. If hearing outcomes were dictated by situational factors, we would expect commissioners to lean into the harsher conditions set by the new law. This means commissioners would skew towards the high end of available denial lengths, based on the broader framing emphasizing the costs to victims as well as the new choice architecture promoting longer denial lengths. When it comes to reasoning, we would expect commissioners to emphasize the harm and degredation the prisoner caused to any victims in their decisions, and spend more time condemning the actions of the prisoner generally to justify longer denials.

Yet if parole outcomes were dictated by commissioners' intuition, and commissioners had stable intuitions built up under the prior regime about what the "right" outcome should be, we would expect Marsy's Law to challenge that intuition. Because commissioners had no prior experience with the new decision choices, the situationist approach would expect the change to mute their moral reasoning. If this were the case, we would expect commissioners to reveal hesitance in their decisions, skewing towards the lower end of denial lengths and providing more subdued moral justifications for those decisions.

In sum, given the changes imposed by Marsy's Law, the situationist approach would expect more severe outcomes and harsher moral reasoning while the intuitionist approach would expect less extreme outcomes and muted reasoning. It is worth noting a third option that we present as a further reference point to guide interpretation of the findings to follow. In a "moral absolutist" view that has traditionally accompanied classical accounts of reasoning since Kant, commissioners would have been arriving at decisions they thought were right in a universal sense, and then map their rational decision onto the available decision options. Meanwhile, the reasons they offer for the decisions would transparently reflect their cognitive deliberation about the decision they reached. If this classical rationalist approach were the case, we would expect everyone who previously would have received a one- or two-year denial to now receive a three-year denial, for example. We would also expect the reasons for denials to be unchanged, since the law left unchanged the reasons commissioners should or could consider for denying parole. Such a rationalist view has been largely abandoned by cognitive scientists (Mercier and Sperber, 2017), yet is a helpful reference point for making sense of our findings.

With these expectations in mind, we now move to describe the data and present basic descriptive findings that motivate our Regression Discontinuity approach.

3 Data and Descriptive Evidence

3.1 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. 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.

3.2 Using NLP to Measure Reasoning

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' reasoning, 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)certainty. 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 the target word.)

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.

3.3 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 4). 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. 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.

3.3.1 Parole Length and Marsy's Law: Decomposition Exercises

We seek to gain further initial descriptive evidence on how Marsy's Law impacted denial lengths through two decomposition exercises. In the post-Law period, the choice of denial lengths increased due to the Law, but how this impacts outcomes was mediated through commissioners' response to the new choice set. The decompositions create a bridge between the pre- and post-Law changes by introducing an intermediate step that holds fixed one of the two dimensions. Recall from Section 2.4 that the situationist framework predicts commissioners would be more severe in their decisions after Marsy's Law due to the framing and anchoring effects of the legal change. We investigate this possibility in the "pure situationist" decomposition. Meanwhile, the rational moral absolutist framework predicts that denial lengths after Marsy's Law would be as close as possible to those before Marsy's Law. We investigate this possibility in the "pure moral absolutist" decomposition. In what follows, the two decomposition exercises provide initial evidence that commissioners were less relatively severe in their decisions than if they had simply applied their relative preferences from before the law to the new denial length options, but more severe than if they had arrived at their decisions as moral absolutists.

3.3.2 Pure Situationist Decomposition

Our first composition asks how a pure situationist archetype would respond to the Law. For this archetype, the important element of the choice set is the ordinal structure of the choice set, rather than the cardinal values of the denial length choices themselves. That is, if purely driven by the choice set, we would expect commissioners to apply their same relative preferences for short, medium, and long denials to the new set of denial length options. This decomposition allows us to compare the observed denial lengths against this expectation.

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 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, i.e., $d_1 = \{0, 3, 5, 7, 10, 15\}$, and again calculate the mean denial length.[3] 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 situationist framework, as it does not take into effect anchoring or broader social signals that would push commissioners to have *more severe* relative preferences after the law. Figure 2 presents this decomposition graphically. Note that as we move from Figure 2a to Figure 2b, the shape of the distribution does not change, just the values on the x-axis. 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.

Third, we calculate denial lengths in the post-Law period. 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 hold fixed the distribution of ordinal choices commissioners make, but translate the values of these ordinal choices into the post-Law choice set. Accordingly, row 2 shows the mean of denial lengths that *would have been* realized had commissioners not changed the ordinal position of their denial length choice. Here, we find that the simulated counterfactual mean denial length would have been 5.3 years. We can contrast this to what we actually find in the post-Law period – a mean of 4.1 years. Figure 2 presents the underlying distribution of denial lengths for this decomposition. We implement a Pearson's χ^2 test to compare the counterfactual and actual 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-specific choice set.

[**Figure 2 here**]

3.3.3 Pure Moral Absolutist Decomposition

The second decomposition that we present captures the response to the Law of a pure moral absolutist archetype. Such a commissioner would not respond to any anchoring, framing, or choice set effects of the new, post-Law denial length options. Rather their behavior would be guided by a rational consideration of the "correct" denial length in an absolute sense, based purely on the facts of the specific case at hand and the commissioners' abstract principles of justice. This means they would choose the denial length that matches on their optimal denial length as closely as possible in both the pre- and post-Law period.

We proceed with the same three steps as the prior decomposition. 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 construct a mapping from pre-Law denial length choice to post-Law choices, implementing the constraint that an absolutist-type commissioner would choose 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. Two sets of choices are noteworthy here. For the choice of 4 years denial 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.

[**Table 3 here**]

Third, and as before, we calculate denial lengths in the post-Law period. We present the results of this decomposition exercise in Table 4, and present the distributions of denial length across the three cases in Figure 3.

[Table 4 here]

The decomposition exercise for the pure rationalist archetype – the results of which can be found in row 2 of Table 4 – highlights a mean denial length that is substantially lower than the actual denial length we document. We once again implement with a Pearson's χ^2 test to compare the counterfactual and actual distributions in the post-Law period, which reveals visually and statistically significantly different distributions.

In summary, the observed decision lengths falls above what we would expect if the commissioners were pure moral absolutists, but below the lower bound of what we would expect if they were purely situationists. This provides initial evidence of commissioners' ambivalence about implementing the law: they do signficantly 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 lean into the more punitive dimensions of the law. Yet to this point, we have only provided a descriptive picture of the law's impacts on outcomes. From here, we introduce our analytic approach for gaining causal traction on this change and investigating how the law impacted the reasons commissioners gave for their decisions.

[Figure 3 here]

4 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 will enable 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 as our treatment variable (implementation of Marsy's Law) 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 at least 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 Marsy's Law 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.

4.1 Empirical Specification

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

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

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-Huber-White standard errors throughout.

4.1.1 Identification

In order to interpret our RDD 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^0_{it}, Y^1_{it}) 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.

5 Regression Discontinuity Design Results

We now present our core RDD results. We start first with parole outcomes. We then move to a series of outcomes based on the transcripts from the hearings. The aim of this analysis is to better understand how parole commissioners explained and justified the on-average longer parole denial lengths they handed down after the enactment of Marsy's Law.

5.1 Marsy's Law and Parole Outcomes

In Table 6 we present RDD 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 raw data (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. To disentangle the effect of the different set of denial length options from how commissioners responded to the different choice sets they faced in the two periods, we normalize denial length by the period-specific range of denial length options.[4] 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 3.3.1: Although commissioners handed down longer denial lengths post-Marsy's Law, 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.

5.2 Marsy's Law and Commissioner Reasoning

We now turn to consider what we can learn about commissioners' responses to the parole mandates of Marsy's Law from the parole hearing transcripts. In particular we are interested in how commissioners rationalize the longer parole denials they hand down post-Mary's Law.

Section 2.4 provided three divergent expectations. Based on changes to the broader context and specific choice architecture, the situationist approach anticipated that commissioners would use more morally charged language and particularly emphasize the harm and degradation to victims. In contrast, the intuitionist approach suggested that the commissioners would be cautious as they sought to adapt their prior practical intuition about decisions to a new environment. This would lead us to expect commissioners to express less overall moral certitude in their decisions. Lastly, the moral absolutist approach would expect commissioners to be unphased by the new Law, and continue to apply their old reasoning unchanged. We present RD estimates for a series of dimensions from the parole hearings in Table 7.

[Table 7 here]

Parole hearings comprise two key components – the proceedings and the decision. A recess when commissioners deliberate separates the two periods. With the exception of columns 1 and 2, all outcomes relate to the decision component of the hearing. 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). In sum, 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; 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 overall 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 of decision 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, commissioners ex-post rationalized their on-average more severe 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 situationist approach 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").[5] However, we do find a shift relative to external reference points – column 9 shows that post-Law, decisions shift 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 situationist framework would predict. 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).

5.3 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).[6] The situationist approach would lead us to expect greater emphasis on harm and degredation, given the emphasis on the victim. Meanwhile, the intuitionist approach would lead us to expect decreasing moral valence, as commissioners would have less strong intuitions about the outcomes.

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.[7] 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. Such underlying heterogeneity is relevant to our work here, as our aim is to document the ex-post rationalization of commissioners to the stricter parole framework mandated by 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 moral pillars 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.

5.4 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 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 few differences across the racial groups. 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 5.2, 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). 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).

Overall, the patterns of differential treatment of Hispanic prisoners serves as an internal validity check for our baseline conclusions – it is for Hispanics that we see the most severe parole outcomes changes post-Law, and it is for Hispanics that we detect the only negative post-Law changes in commissioners' moral valence. This fits with an intuitionist framework, where commissioners would have a stronger intuition about the hearing outcome for Hispanic prisoners and thus reach for stronger moral language to justify that outcome. Meanwhile, the situationist framework does not offer a parsimonious explanation of this pattern, as one would have to explain why the changes to the choice architecture itself would trigger such changes for only prisoners of one racial group.

6 Conclusion

What is the relationship between reason-giving and decision-making in legal bureaucratic contexts? In this work, we make use of rich textual data from transcripts of administrative decisions and leverage a voter-mandated change to the decision-making context to gain new traction on this question. Specifically, we examine the California Parole Board's decisions about whether to release someone from prison, and if denied, how long before that person becomes eligible for another hearing. In 2008 California voters passed Marsy's Law, which led to a much more severe set of denial length options for parole commissioners and elevated victims to greater symbolic salience in the proceedings. The new law left unchanged the substantive factors that commissioners were to consider in reaching their decisions.

Recent advances in understanding practical decision-making and reason-giving have moved us beyond classical conceptions of reasoning as transparent and logical (Mercier and Sperber, 2017) – a view which would predict commissioners' reasoning to be unmoved by this change. Recently, sociologists have called for integrating new models of reasoning into our sociological accounts (Bruch and Feinberg, 2017; Vaisey and Valentino, 2018). Yet alternate models of reasoning provide competing expectations for how commissioners would respond to this law. What we call the *situationist* approach emphasizes the role of context in shaping people's decisions, and the reasons they give for those decisions (Jerolmack and Khan, 2014; Mills, 1940; Ross and Nisbett, 1991; Thaler, 2016). The changes the law made to the "choice architecture" of decisions to be more punitive, coupled with signaling about the sanctity of the victim, implies that the new law should have led commissioners to be more severe in their decisions and more morally condemning in their explanations. However, what we label the *intuitionist* approach emphasizes the stability of individual's intuitions in guiding decisions, and argues that reason-giving follows from individual intuition (Haidt, 2012; Mercier and Sperber, 2017; Vaisey, 2009). This approach implies that the law would have destabilized commissioners' intuitions about the right decision to reach, leading to relatively less severe outcomes and more muted moral reasoning in their explanations.

We use Natural Language Processing methods and a Regression Discontinuity design to examine how commissioners responded to this shift both in the substantive outcomes they arrived at, and the reasons they gave for those outcomes. 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 their new choice set. This behavioral response tempered the mechanical denial length-increasing effect of the Law – the law increased all denial options by a magnitude of three, and the situationist approach would have expected commissioners' decisions to increase by at least this much. 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. Combining our findings on commissioner decision-making with commissioner reasoning suggests unease with the more punitive mandates of the Law. However, the consequences were not the same for all prisoners: in particular, for Hispanic people appearing before the board, commissioners gave longer relative denial lengths and invoked more negative moral language relative to Black and white people. Together, the findings fit well within an intuitionist view of reason that suggests the new law destabilized commissioners' instincts about what the "correct" decision should be. Meanwhile, the findings cut against the expectations of a pure situationist view of reason which would predict commissioners to lean into the punitive and condemnatory thrust of the new law. Ironically, then, the law appears to have made the hearings *more* bureaucratic – with shorter decisions, more invocation of procedure and rules, and less morally strident language.

Substantively, our findings are of import for understanding decision-making under legal reforms. Our case and design allow us to make progress on understanding *how* decision-makers make decisions. In line with the intuitionist approach, our evidence suggests that commissioners' established intuitions, built up over prior hearings under the old legal framework, made them less clear about what the "right" outcome should be under the new law. The findings points us towards intuition of decision-makers in understanding the consequences of legal reform on administrative procedures more broadly. Simply put, prior practical experience and training – *who* is making the decisions – matters. A focus on simply the rules and context of decision-making, without consideration of the decision-makers and their experience, backgrounds, and training, will fail to anticipate how a reform plays out in practice. At the same time, while commissioners mediated the impact of the law, they did not negate it. 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.

Methodologically, our study highlights the benefits of a text-as-data approach to the setting of legal decision-making and the possibilities of using such approaches within a causal inference framework. The allure of NLP-based methods when applied to text where legal decision-makers are required to explain their decisions is that it allows us to document previously invisible linkages between the actions and the explanations of legal decision-makers. As many have noted, there is not a one-to-one correspondence between action and explanation (Jerolmack and Khan, 2014). Yet rather than presenting a methodological pickle, text-as-data methods provide novel opportunities to expand our understanding of the relationship between talk and action in cases where we can clearly observe both.

Notes

[1]Unlike 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.

[2]See the following 12/16/08 press release: https://web.archive.org/web/20081218130340/http://www.cdcr.ca.gov/News/2008_Press_Releases/Dec_16.html.

[3]Mechanically, 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.

[4]For 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).

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

[6]To 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.

[7]We list these in Table A1.

References

Jisun An, Haewoon Kwak, and Yong-Yeol Ahn. SemAxis: A Lightweight Framework to Characterize Domain-Specific Word Semantics Beyond Sentiment. In Iryna Gurevych and Yusuke Miyao, editors, *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 2450–2461, Melbourne, Australia, July 2018. Association for Computational Linguistics. doi: 10.18653/v1/P18-1228. URL https://aclanthology.org/P18-1228.

Maria Antoniak and David Mimno. Evaluating the Stability of Embedding-based Word Similarities. *Transactions of the Association for Computational Linguistics*, 6:107–119, 2018. doi: 10.1162/tacl_a_00008. URL https://www.aclweb.org/anthology/Q18-1008.

David Arnold, Will Dobbie, and Crystal S Yang. Racial bias in bail decisions. *The Quarterly Journal of Economics*, 133(4):1885–1932, 05 2018.

David Arnold, Will Dobbie, and Peter Hull. Measuring racial discrimination in bail decisions. *American Economic Review*, 112(9):2992–3038, September 2022.

Elliott Ash and Stephen Hansen. Text algorithms in economics. *Annual Review of Economics*, 15(Volume 15, 2023):659–688, 2023.

Elliott Ash, Daniel L. Chen, and Suresh Naidu. Ideas have Consequences: The Impact of Law and Economics on American Justice. *Working Paper*, 2018. URL http://elliottash.com/wp-content/uploads/2018/08/ash-chen-naidu-2018-07-15.pdf.

Elliott Ash, Daniel L. Chen, and Sergio Galletta. Measuring Judicial Sentiment: Methods and Application to US Circuit Courts. *Economica*, 89(354): 362–376, 2022. ISSN 1468-0335. doi: 10.1111/ecca.12397. URL https://onlinelibrary.wiley.com/doi/abs/10.1111/ecca.12397. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/ecca.12397.

Hadar Aviram. *Yesterday's Monsters: The Manson Family Cases and the Illusion of Parole*. University of California Press, Oakland, CA, 2020.

Rachel Kahn Best and Alina Arseniev-Koehler. The Stigma of Diseases: Unequal Burden, Uneven Decline. *American Sociological Review*, 88(5):938–969, October 2023. ISSN 0003-1224. doi: 10.1177/00031224231197436. URL https://doi.org/10.1177/00031224231197436. Publisher: SAGE Publications Inc.

Manudeep Bhuller, Gordon B. Dahl, Katrine V. Løken, and Magne Mogstad. Incarceration, recidivism, and employment. *Journal of Political Economy*, 128(4):1269–1324, 2020.

Alan S. Blinder. Wage discrimination: Reduced form and structural estimates. *The Journal of Human Resources*, 8(4):436–455, 1973.

Pierre Bourdieu. *The Logic of Practice*. Stanford University Press, Stanford, CA, 1990.

Elizabeth Bruch and Fred Feinberg. Decision-Making Processes in Social Contexts. *Annual Review of Sociology*, 43(1):207–227, 2017. doi: 10.1146/annurev-soc-060116-053622. URL `https://doi.org/10.1146/annurev-soc-060116-053622`. _eprint: https://doi.org/10.1146/annurev-soc-060116-053622.

Sebastian Calonico, Matias D. Cattaneo, and Rocio Titiunik. Robust nonparametric confidence intervals for regression-discontinuity designs. *Econometrica*, 82(6):2295–2326, 2014.

Gordon B. Dahl, Andreas Ravndal Kostøl, and Magne Mogstad. Family Welfare Cultures. *The Quarterly Journal of Economics*, 129(4):1711–1752, 08 2014.

Isaac Dalke. I Come before You a Changed Man: "Insight," Compliance, and Refurbishing Penal Practice in California. *Law & Social Inquiry*, 49(2):1138–1168, May 2024. ISSN 0897-6546, 1747-4469. doi: 10.1017/lsi.2023.20. URL `https://www.cambridge.org/core/journals/law-and-social-inquiry/article/i-come-before-you-a-changed-man-insight-compliance-and-refurbishing-penal-practice-in-california/C08B2D845DB27DB46DF1DEEFD2C656F8`.

Shai Danziger, Jonathan Levav, and Liora Avnaim-Pesso. Extraneous factors in judicial decisions. *Proceedings of the National Academy of Sciences*, 108(17):6889–6892, April 2011. doi: 10.1073/pnas.1018033108. URL `https://www.pnas.org/doi/abs/10.1073/pnas.1018033108`. Publisher: Proceedings of the National Academy of Sciences.

Will Dobbie, Jacob Goldin, and Crystal S. Yang. The effects of pretrial detention on conviction, future crime, and employment: Evidence from randomly assigned judges. *American Economic Review*, 108(2):201–40, February 2018.

Emile Durkheim. *Durkheim and the Law*. Palgrave Macmillan, Basingstoke, UK, second edition edition, 2013.

Birte Englich. Blind or Biased? Justitia's Susceptibility to Anchoring Effects in the Courtroom Based on Given Numerical Representations. *Law & Policy*, 28(4):497–514, 2006. ISSN 1467-9930. doi: 10.1111/j.1467-9930.2006.00236.x. URL `https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1467-9930.2006.00236.x`. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1467-9930.2006.00236.x.

Benjamin Enke. Moral Values and Voting. *Journal of Political Economy*, 128(10):3679–3729, October 2020. ISSN 0022-3808. doi: 10.1086/708857. URL `https://www.journa` ls.uchicago.edu/doi/full/10.1086/708857. Publisher: The University of Chicago Press.

Charles R. Epp, Steven Maynard-Moody, and Donald Haider-Markel. *Pulled Over: How Police Stops Define Race and Citizenship.* University of Chicago Press, April 2014. ISBN 978-0-226-11404-0. Google-Books-ID: TlHmAgAAQBAJ.

Amir Feder, 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. Causal inference in natural language processing: Estimation, prediction, interpretation and beyond. *Transactions of the Association for Computational Linguistics*, 10:1138–1158, 2022.

Brigham R Frandsen. 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*, pages 281–315. Emerald Publishing Limited, 2017.

J.A. Frimer, R. Boghrati, J. Graham, and M. Dehgani. Moral Foundations Dictionary for Linguistic Analyses 2.0. April 2019. URL https://osf.io/https://osf.io/xakyw.

Nikhil Garg, Londa Schiebinger, Dan Jurafsky, and James Zou. Word embeddings quantify 100 years of gender and ethnic stereotypes. *Proceedings of the National Academy of Sciences*, 115(16):E3635–E3644, April 2018. ISSN 0027-8424, 1091-6490. doi: 10.1073/pnas.1720347115. URL https://www.pnas.org/content/115/16/E3635.

Matthew Gentzkow, Bryan Kelly, and Matt Taddy. Text as data. *Journal of Economic Literature*, 57(3):535–74, September 2019.

Yoav Goldberg and Omer Levy. word2vec Explained: deriving Mikolov et al.'s negative-sampling word-embedding method. *arXiv:1402.3722 [cs, stat]*, February 2014. URL http://arxiv.org/abs/1402.3722. arXiv: 1402.3722.

Philip Goodman. "It's Just Black, White, or Hispanic": An Observational Study of Racializing Moves in California's Segregated Prison Reception Centers. *Law & Society Review*, 42(4):735–770, December 2008. ISSN 1540-5893. doi: 10.1111/j.1540-5893.2008.00357.x. URL https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1540-5893.2008.00357.x.

Jesse Graham, Jonathan Haidt, Sena Koleva, Matt Motyl, Ravi Iyer, Sean P. Wojcik, and Peter H. Ditto. Chapter Two - Moral Foundations Theory: The Pragmatic Validity of Moral Pluralism. In Patricia Devine and Ashby Plant, editors, *Advances in Experimental Social Psychology*, volume 47, pages 55–130. Academic

Press, January 2013. doi: 10.1016/B978-0-12-407236-7.00002-4. URL https://www.sciencedirect.com/science/article/pii/B9780124072367000024.

Joss Greene and Isaac Dalke. "You're still an angry man": Parole boards and logics of criminalized masculinity. *Theoretical Criminology*, 25(4):639–662, November 2021. ISSN 1362-4806. doi: 10.1177/1362480620910222. URL https://doi.org/10.1177/1362480620910222. Publisher: SAGE Publications Ltd.

Jonathan Haidt. *The Righteous Mind: Why Good People are Divided by Politics and Religion.* Vintage Books, New York, NY, 2012.

David J. Harding, Jeffrey D. Morenoff, Anh P. Nguyen, Shawn D. Bushway, and Ingrid A. Binswanger. A natural experiment study of the effects of imprisonment on violence in the community. *Nature Human Behaviour*, 3(7):671–677, July 2019. ISSN 2397-3374. doi: 10.1038/s41562-019-0604-8. URL https://www-nature-com.libproxy.berkeley.edu/articles/s41562-019-0604-8. Publisher: Nature Publishing Group.

Carrie L. Hempel. Lawrence and Shaputis and Their Impact on Parole Decisions in California Probation & Parole Reform. *Federal Sentencing Reporter*, 22:176–180, 2010.

Steve Herbert. Degradation or Redemption? A Parole Board Polices a Moral Boundary. *Law & Social Inquiry*, pages 1–21, December 2022. ISSN 0897-6546, 1747-4469. doi: 10.1017/lsi.2022.27. URL https://www.cambridge.org/core/journals/law-and-social-inquiry/article/degradation-or-redemption-a-parole-board-polices-a-moral-boundary/92E802756079F7D77C6721A6BD9296DA. Publisher: Cambridge University Press.

Steven Hitlin and Stephen Vaisey. The New Sociology of Morality. *Annual Review of Sociology*, 39(Volume 39, 2013):51–68, July 2013. ISSN 0360-0572, 1545-2115. doi: 10.1146/annurev-soc-071312-145628. URL https://www.annualreviews.org/content/journals/10.1146/annurev-soc-071312-145628. Publisher: Annual Reviews.

Christopher K. Hsee. Value seeking and prediction-decision inconsistency: Why don't people take what they predict they'll like the most? *Psychonomic Bulletin & Review*, 6(4):555–561, December 1999. ISSN 1531-5320. doi: 10.3758/BF03212963. URL https://doi.org/10.3758/BF03212963.

Colin Jerolmack and Shamus Khan. Talk Is Cheap: Ethnography and the Attitudinal Fallacy. *Sociological Methods & Research*, March 2014. doi: 10.1177/0049124114523396. URL https://journals.sagepub.com/doi/10.1177/0049124114523396.

Eric J. Johnson, 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. Beyond nudges: Tools of a choice architecture. *Marketing Letters*, 23(2):487–504, June 2012. ISSN 1573-059X. doi: 10.1007/s11002-012-9186-1. URL https://doi.org/10.1007/s11002-012-9186-1.

Daniel Kahneman and Amos Tversky. Choices, values, and frames. *American Psychologist*, 39(4):341–350, 1984. ISSN 1935-990X. doi: 10.1037/0003-066X.39.4.341. Place: US Publisher: American Psychological Association.

Evelyn M. Kitagawa. Components of a difference between two rates. *Journal of the American Statistical Association*, 50(272):1168–1194, 1955.

Jeffrey R. Kling. Incarceration length, employment, and earnings. *American Economic Review*, 96(3):863–876, June 2006.

Austin C. Kozlowski, Matt Taddy, and James A. Evans. The Geometry of Culture: Analyzing the Meanings of Class through Word Embeddings. *American Sociological Review*, 84(5):905–948, September 2019. ISSN 0003-1224. doi: 10.1177/0003122419877135. URL https://doi.org/10.1177/0003122419877135.

Michael Lipsky. *Street-Level Bureaucracy: Dilemmas of the Individual in Public Service*. Russell Sage Foundation, New York, NY, 30th anniversary edition edition, 2010.

Omar Lizardo and Michael Strand. Skills, toolkits, contexts and institutions: Clarifying the relationship between different approaches to cognition in cultural sociology. *Poetics*, 38(2):205–228, April 2010. ISSN 0304-422X. doi: 10.1016/j.poetic.2009.11.003. URL https://www.sciencedirect.com/science/article/pii/S0304422X09000606.

John Levi Martin. *The Explanation of Social Action*. Oxford University Press, Oxford, UK, 2011.

Julia Mendelsohn, Yulia Tsvetkov, and Dan Jurafsky. A Framework for the Computational Linguistic Analysis of Dehumanization. *Frontiers in Artificial Intelligence*, 3, 2020. ISSN 2624-8212. URL https://www.frontiersin.org/articles/10.3389/frai.2020.00055.

Hugo Mercier and Dan Sperber. *The Enigma of Reason*. Harvard University Press, Cambridge, MA, 2017.

Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. Efficient Estimation of Word Representations in Vector Space. *arXiv:1301.3781 [cs]*, January 2013. URL http://arxiv.org/abs/1301.3781. arXiv: 1301.3781.

C. Wright Mills. Situated Actions and Vocabularies of Motive. *American Sociological Review*, 5(6):904–913, 1940. ISSN 0003-1224. doi: 10.2307/2084524. URL http://www.jstor.org/stable/2084524.

Saif Mohammad. Obtaining Reliable Human Ratings of Valence, Arousal, and Dominance for 20,000 English Words. In Iryna Gurevych and Yusuke Miyao, editors, *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 174–184, Melbourne, Australia, July 2018. Association for Computational Linguistics. doi: 10.18653/v1/P18-1017. URL `https://aclanthology.org/P18-1017`.

Jerzy Neyman. On the application of probability theory to agricultural experiments. essay on principles. *Ann. Agricultural Sciences*, pages 1–51, 1923.

Ronald Oaxaca. Male-female wage differentials in urban labor markets. *International Economic Review*, 14(3):693–709, 1973.

Anna Offit. Race-Conscious Jury Selection. *Ohio State Law Journal*, 82(2):201–248, 2021. URL `https://heinonline.org/HOL/P?h=hein.journals/ohslj82&i=209`.

Derek L. Phillips and Kevin J. Clancy. Some Effects of "Social Desirability" in Survey Studies. *American Journal of Sociology*, 77(5):921–940, 1972. ISSN 0002-9602. URL `https://www.jstor.org/stable/2776929`. Publisher: University of Chicago Press.

Marie Pryor, Kim Shayo Buchanan, and Phillip Atiba Goff. Risky Situations: Sources of Racial Disparity in Police Behavior. *Annual Review of Law and Social Science*, 16 (Volume 16, 2020):343–360, October 2020. ISSN 1550-3585, 1550-3631. doi: 10.1146/annurev-lawsocsci-101518-042633. URL `https://www.annualreviews.org/content/journals/10.1146/annurev-lawsocsci-101518-042633`. Publisher: Annual Reviews.

Jeffrey Rachlinski and Andrew Wistrich. Gains, Losses, and Judges: Framing and the Judiciary. *Cornell Law Faculty Publications*, December 2018. URL `https://scholarship.law.cornell.edu/facpub/1703`.

Matthew D Rocklage, Sharlene He, Derek D Rucker, and Loran F Nordgren. Beyond sentiment: The value and measurement of consumer certainty in language. *Journal of Marketing Research*, 60(5):870–888, 2023.

Emma Rodman. A Timely Intervention: Tracking the Changing Meanings of Political Concepts with Word Vectors. *Political Analysis*, 28(1):87–111, 2020. doi: 10.1017/pan.2019.23.

Pedro L. Rodriguez and Arthur Spirling. Word Embeddings: What Works, What Doesn't, and How to Tell the Difference for Applied Research. *The Journal of Politics*, 84 (1):101–115, January 2022. ISSN 0022-3816. doi: 10.1086/715162. URL `https://www.journals.uchicago.edu/doi/full/10.1086/715162`. Publisher: The University of Chicago Press.

Lee Ross and Richard E. Nisbett. *The person and the situation: perspectives of social psychology / Lee Ross, Richard E. Nisbett.* Temple University Press, Philadelphia, 1991. ISBN 978-0-87722-851-6.

Donald B Rubin. Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of educational Psychology*, 66(5):688, 1974.

Simone Schnall, Jonathan Haidt, Gerald L. Clore, and Alexander H. Jordan. Disgust as Embodied Moral Judgment. *Personality and Social Psychology Bulletin*, 34(8):1096–1109, August 2008. ISSN 0146-1672. doi: 10.1177/0146167208317771. URL https://doi.org/10.1177/0146167208317771. Publisher: SAGE Publications Inc.

Victor L Shammas. The perils of parole hearings: California lifers, performative disadvantage and the ideology of insight. *Political and Legal Anthropology Review*, Forthcoming, 2018.

Philip Smith. *Punishment and Culture*. University of Chicago Press, Chicago, IL, 2008.

Richard H. Thaler. Behavioral Economics: Past, Present, and Future. *American Economic Review*, 106(7):1577–1600, July 2016. ISSN 0002-8282. doi: 10.1257/aer.106.7.1577. URL https://www.aeaweb.org/articles?id=10.1257/aer.106.7.1577.

Richard H. Thaler, Cass R. Sunstein, and John P. Balz. Chapter 25. Choice Architecture. In *Chapter 25. Choice Architecture*, pages 428–439. Princeton University Press, January 2013. ISBN 978-1-4008-4534-7. doi: 10.1515/9781400845347-029. URL https://www.degruyter.com/document/doi/10.1515/9781400845347-029/pdf?licenseType=restricted.

Amos Tversky and Daniel Kahneman. Judgment under Uncertainty: Heuristics and Biases. *Science*, 185(4157):1124–1131, September 1974. doi: 10.1126/science.185.4157.1124. URL https://www.science.org/doi/10.1126/science.185.4157.1124. Publisher: American Association for the Advancement of Science.

Stephen Vaisey. Motivation and Justification: A Dual-Process Model of Culture in Action. *American Journal of Sociology*, 114(6):1675–1715, May 2009. ISSN 0002-9602. doi: 10.1086/597179. URL https://www.journals.uchicago.edu/doi/10.1086/597179.

Stephen Vaisey and Lauren Valentino. Culture and choice: Toward integrating cultural sociology with the judgment and decision-making sciences. *Poetics*, 68:131–143, June 2018. ISSN 0304-422X. doi: 10.1016/j.poetic.2018.03.002. URL https://www.sciencedirect.com/science/article/pii/S0304422X17301869.

Keith Wattley. Insight into California's Life Sentences. *Federal Sentencing Reporter*, 25(4):1–5, 2013.

Max Weber. Bureaucracy. In H.H. Gerth and C. Wright Mills, editors, *From Max Weber: Essays in Sociology*, pages 196–244. Oxford University Press, New York, NY, 1946.

Robert Weisberg, Debbie Mukamal, and Jordan Segall. 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, 2011.

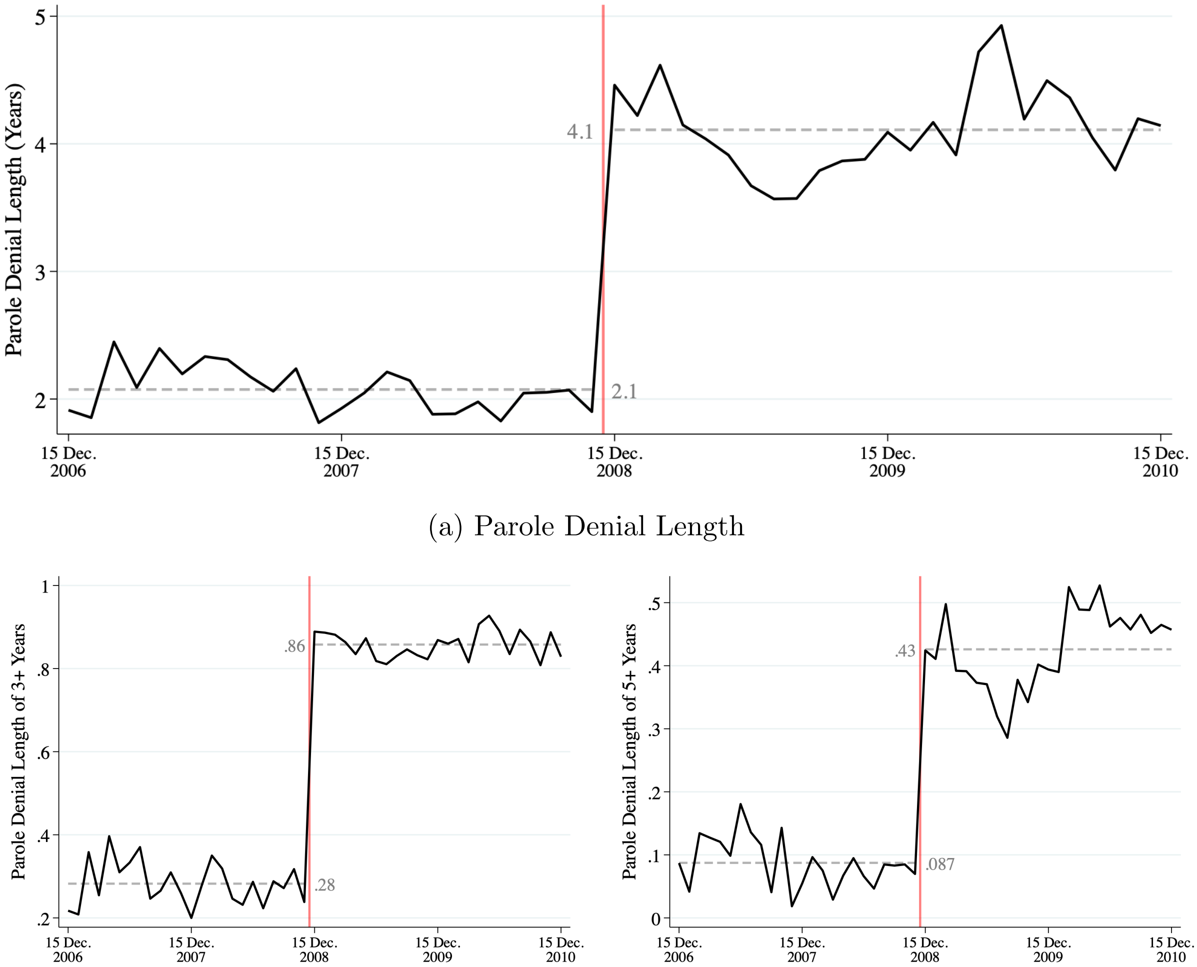
Timothy D. Wilson, Douglas J. Lisle, Jonathan W. Schooler, Sara D. Hodges, Kristen J. Klaaren, and Suzanne J. LaFleur. Introspecting about Reasons can Reduce Post-Choice Satisfaction. *Personality and Social Psychology Bulletin*, 19(3):331–339, June 1993. ISSN 0146-1672. doi: 10.1177/0146167293193010. URL https://doi.org/10.1177/0146167293193010. Publisher: SAGE Publications Inc.

Kathryne Young, Debbie Mukamal, and Thomas Favre-Bulle. Predicting Parole Grants: An Analysis of Suitability Hearings for California's Lifer Inmates. *Federal Sentencing Reporter*, 28(4):268–277, 2016.

Kathryne M. Young. Parole Hearings and Victims' Rights: Implementation, Ambiguity, and Reform. *Connecticut Law Review*, 49:431, 2016. URL https://heinonline.org/HOL/Page?handle=hein.journals/conlr49&id=449&div=&collection=.

Figures

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



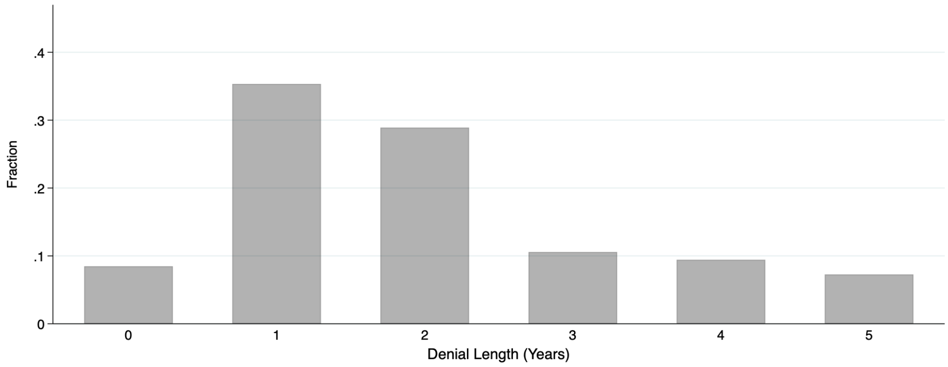
(a) Parole Denial Length

(b) Parole Denial: 3+ Years

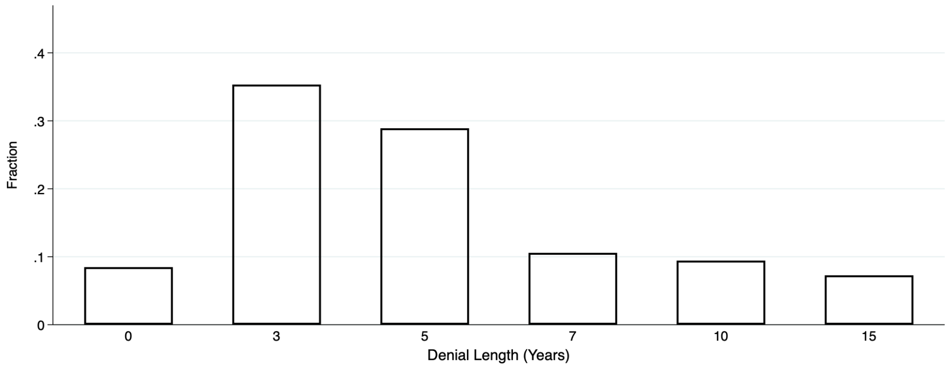
(c) Parole Denial: 5+ Years

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.

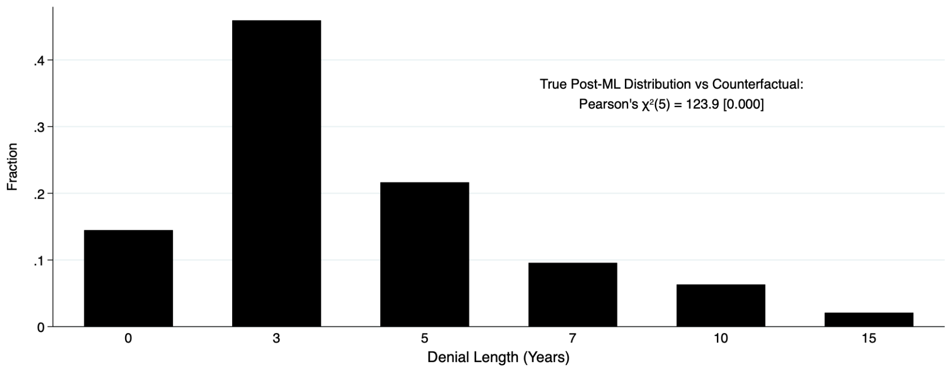
Figure 2: The Distribution of Denial Length Under Pure Behavioralist Decomposition



(a) Pre-Marsy's Law Denial Length, Pre-Mary's Law Choice Set



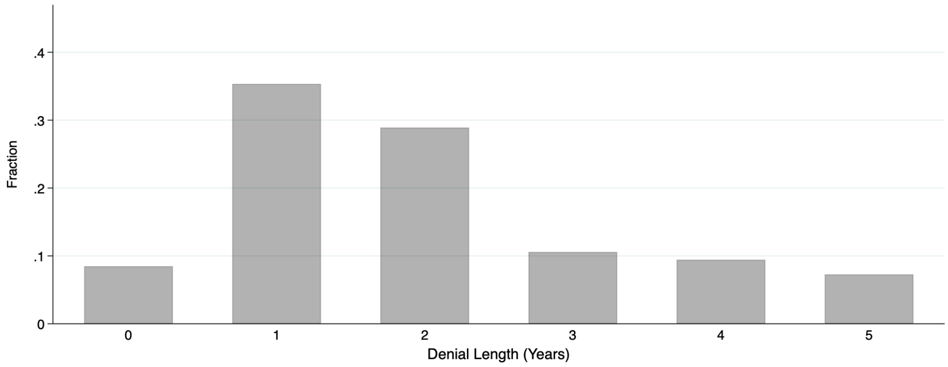
(b) Pre-Marsy's Law Denial Length, Post-Mary's Law Choice Set



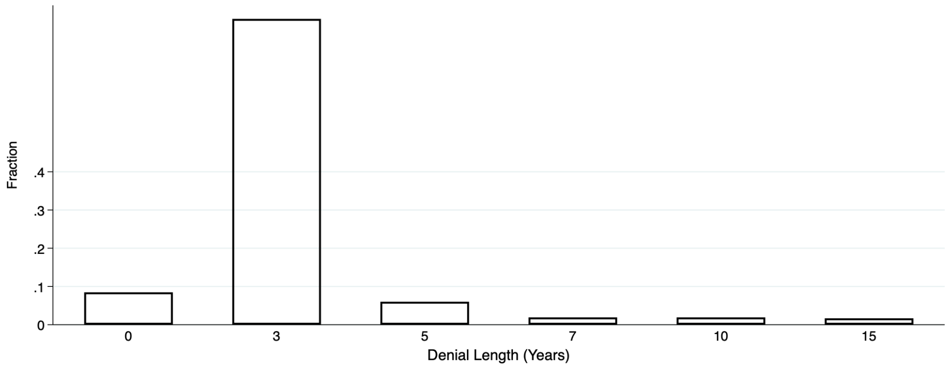
(c) Post-Marsy's Law Denial Length, Post-Mary's Law Choice Set

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 – the counterfactual simulation – for the parole hearings in the pre-Law period, we map actual denial lengths to their ordinal counterparts in the post-Law period. 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, and denote these transformed denial lengths as simulated denial lengths. We represent these in gray. On the same graph, we present the realized, post-Law denial length distribution of denial lengths in black.

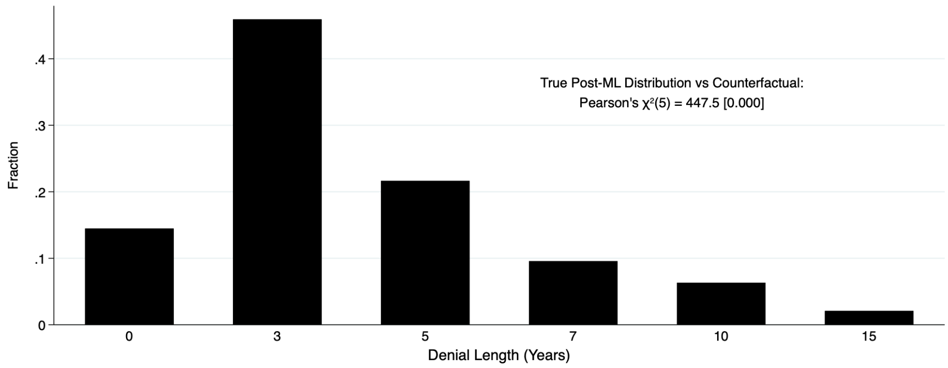
Figure 3: The Distribution of Denial Length Under Pure Rationalist Decomposition



(a) Pre-Marsy's Law Denial Length, Pre-Mary's Law Choice Set



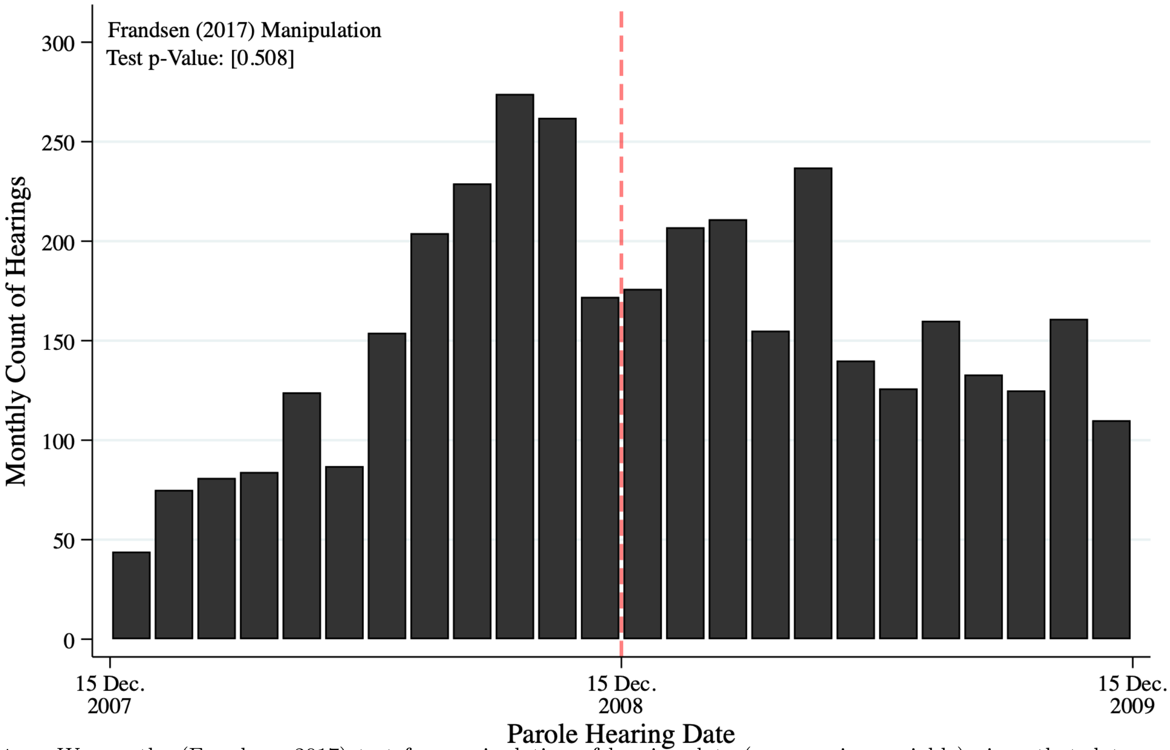
(b) Pre-Marsy's Law Denial Length, Post-Mary's Law Choice Set



(c) Post-Marsy's Law Denial Length, Post-Mary's Law Choice Set

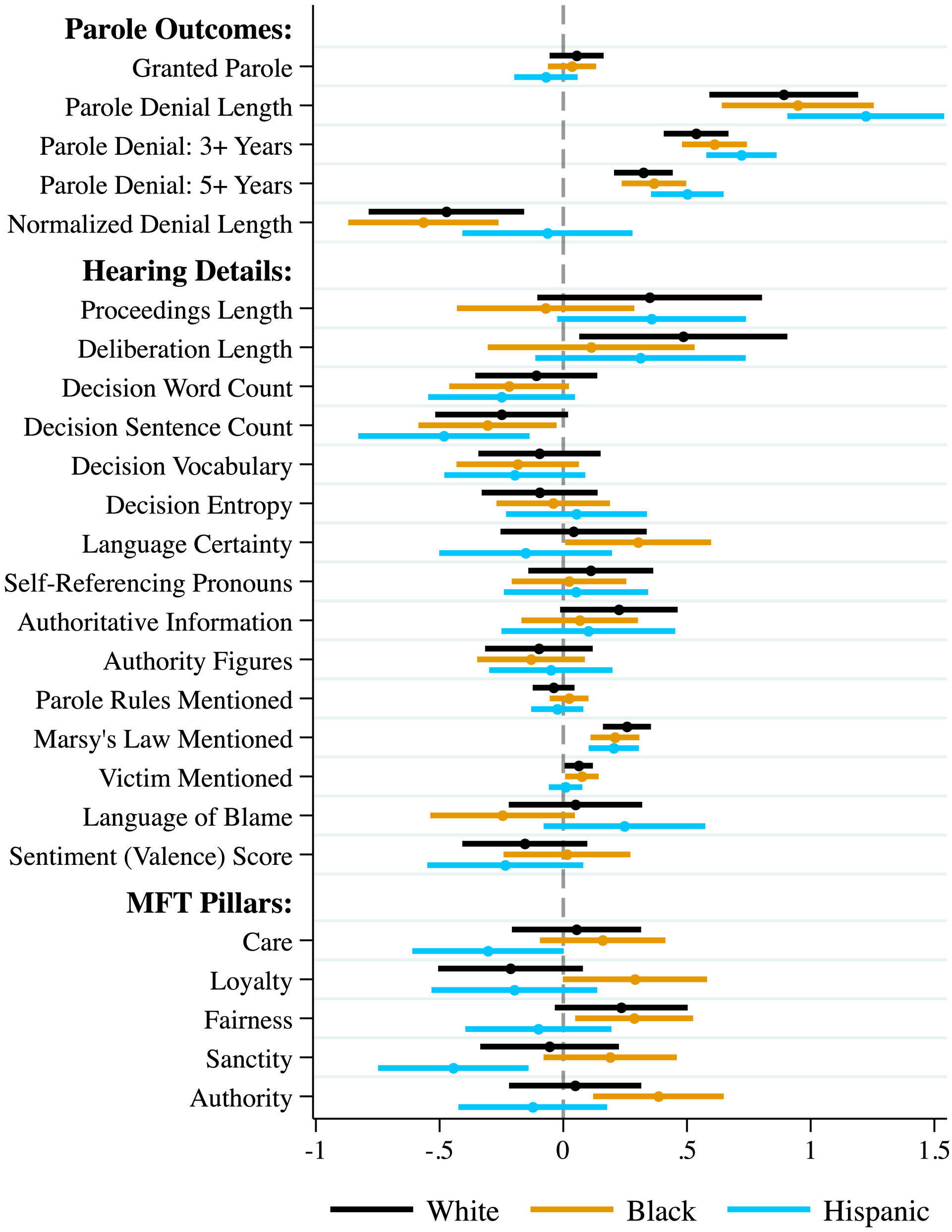
Notes: The figure is based on parole hearings within a 295 day window around the introduction of Marsy's Law in California. In Figure 2b – the counterfactual simulation – for the parole hearings in the pre-Law period, we map actual denial lengths to their ordinal counterparts in the post-Law period. 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, and denote these transformed denial lengths as simulated denial lengths. We represent these in gray. On the same graph, we present the realized, post-Law denial length distribution of denial lengths in black.

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-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.

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	63.5
	(9.32)	(9.3)
Time Served (Days)	7,716	7,741
	(2,357)	(2,325)
Most Severe Offense: Murder	.769	.768
Most Severe Offense: Sex Crime	.0363	.0349
Granted Parole	.11	.114
Denial Length (Years)	3.17	3.02
	(2.51)	(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 Behavioralist Response

	(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 behavioralist response to the Law.

Table 3: Denial Length Decomposition Mapping for the Pure Rationalist Response

	(1)	(2)
Original Denial Length	**Simulated Denial Length**	**Probabilty 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 rationalist 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 Rationalist Response

	(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 rationalist response 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	.0218	-.0404	-.0806	.385	354	.0271	.0274
	(.0477)	(.0436)	(.0351)	(.0509)	(.909)	(225)	(.0424)	(.0197)
$\overline{Y}_{PRE}$	.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-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 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	2.59***	.618***	.396***	-.0912***
	(.0313)	(.228)	(.0388)	(.0379)	(.0227)
$\overline{Y}_0$	.0847	1.99	.273	.0727	.398
Marsy's Law/ $\overline{Y}_0$	.0866	1.3***	2.27***	5.44***	-.229***
	(.369)	(.115)	(.142)	(.521)	(.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) **Proceedings Length (Minutes)**	(2) **Deliberation Length (Minutes)**	(3) **Decision Word Count**	(4) **Decision Sentence Count**	(5) **Decision Vocabulary Size**
Marsy's Law	6.31	5.65**	–152**	–14.1***	–25.2**
	(4.21)	(2.2)	(63.6)	(3.64)	(11.5)
$\overline{Y}_0$	88.9	34	2085	104	606
Observations	2,390	1,842	3,293	3,293	3,293
	(6) **Decision Entropy Score**	(7) **Language Certainty Score**	(8) **Self-Referencing Pronouns**	(9) **Authoritative Information & Documentation**	(10) **Authority Figures**
Marsy's Law	-.0092	.0141	.0696	.129*	-.114*
	(.0103)	(.0174)	(.073)	(.0726)	(.0685)
$\overline{Y}_0$	5.43	6.28	-.0416	-.0361	.141
Observations	3,293	3,293	3,293	3,293	3,293
	(11) **Parole Rules Mentioned**	(12) **Marsy's Law Mentioned**	(13) **Victim Mentioned**	(14) **Language of Blame**	(15) **Sentiment (Valence) Score**
Marsy's Law	-.0101	.224***	.0584***	-.0335	-.00205
	(.0249)	(.0287)	(.0187)	(.0838)	(.00198)
$\overline{Y}_0$	.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-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 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	-.003	.173**	-.0344	.127*
	(.0746)	(.08)	(.0719)	(.0794)	(.0762)
(b) Positive Axis Only					
Marsy's Law	-.118	-.211***	.0464	-.241***	-.0788
	(.0809)	(.0725)	(.0817)	(.08)	(.0704)
(c) Negative Axis Only					
Marsy's Law	-.0824	-.173**	-.121*	-.156**	-.122*
	(.0733)	(.0743)	(.0731)	(.0723)	(.0695)
(d) Top 10 Words Only					
Marsy's Law	-.0127	.104	.173**	-.153*	.27***
	(.0708)	(.0785)	(.0701)	(.0796)	(.0779)
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 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***	.0488	.0507	-.194	.221*	-.0538	.0454
	(.0393)	(.136)	(.127)	(.136)	(.129)	(.139)	(.129)
Black Inmates							
Marsy's Law	-.139***	-.243	.151	.264*	.27**	.184	.364***
	(.038)	(.148)	(.123)	(.136)	(.115)	(.133)	(.128)
Hispanic Inmates							
Marsy's Law	-.0157	.245	-.289*	-.18	-.0948	-.431***	-.117
	(.043)	(.165)	(.149)	(.156)	(.142)	(.151)	(.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

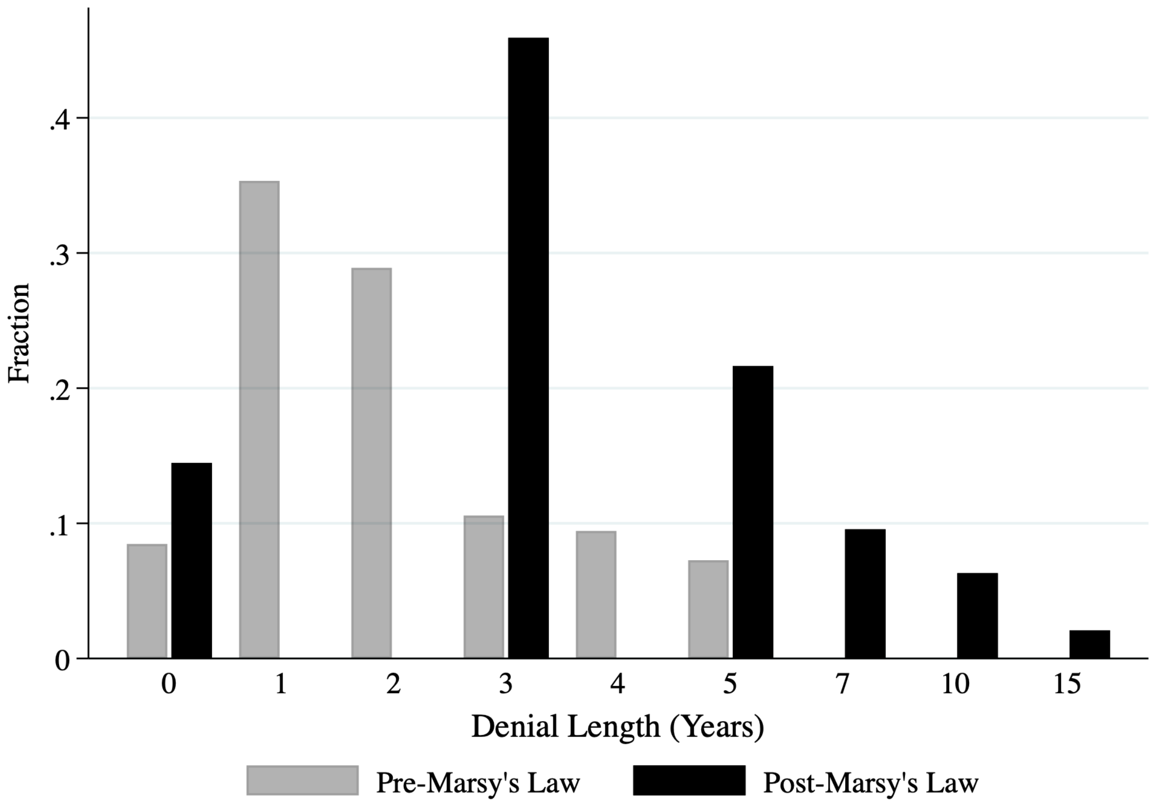
Notes: *** denotes significance at 1%, ** at 5%, and * at 10%. Eicker-Huber-White standard errors in parentheses. Marsy's Law is an indicator variable that takes the value of 1 for hearings from 15 December 2008 and 0 otherwise. Given the temporal structure of commissioner visits to prison, we include 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.

Appendix

A Additional Results

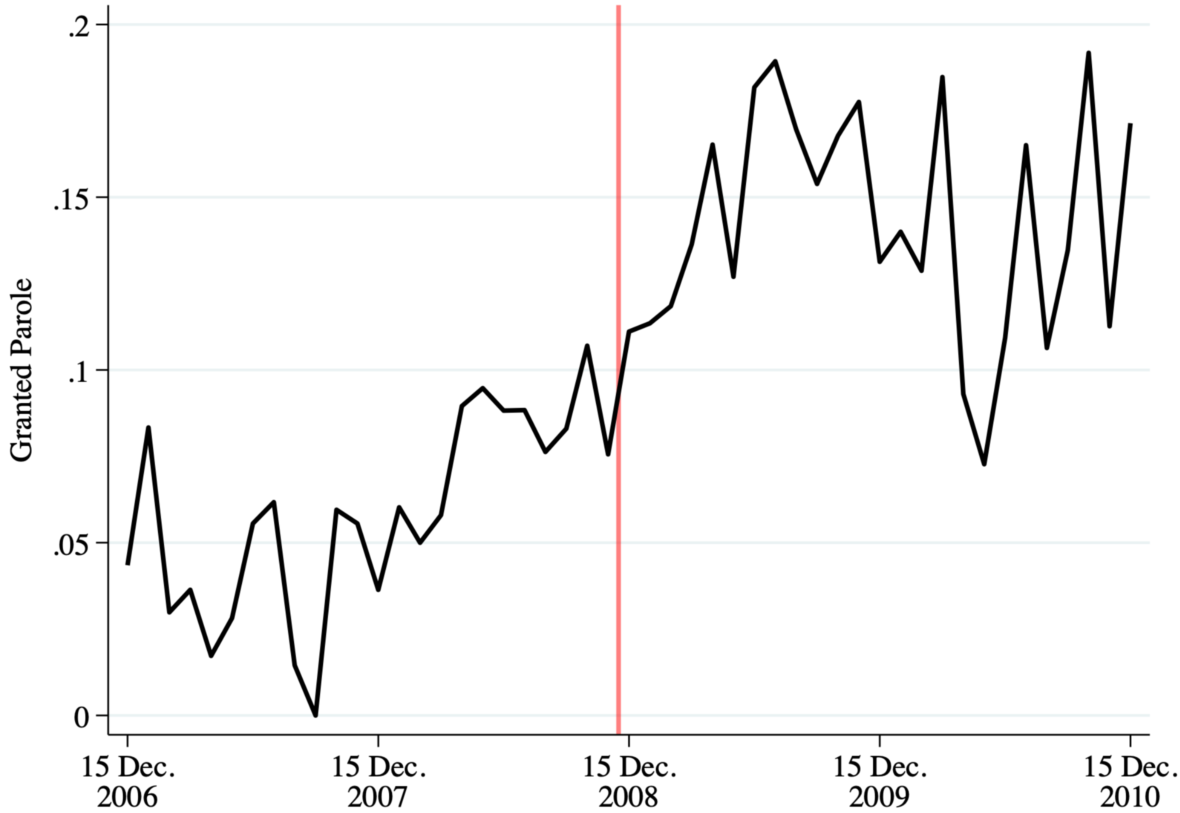
A.1 Denial Length Distribution and the Law

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



A.2 Grant Rate Around Marsy's Law

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.

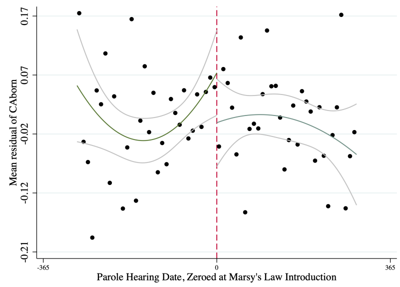
A.3 RDD Continuity Assumption: Graphical Evidence

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

(a) Black

(b) Hispanic

(c) Born in US



(d) Born in CA

(e) Age

(f) Time Served

(g) Murder

(h) Sex Crimes

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 are the bases of our NLP semantic axes output – the MFT pillar axes and the axis for the language of blame.

In Table A2, we present an extended version of the sensitivity analyses in Table 8, splitting the sample by commissioner severity.

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	-.053	-.0391	.00871	-.0372	-.111	.0162	-.0157	.0558	.0261	.389***	.0429	.22**
	(.12)	(.108)	(.115)	(.106)	(.116)	(.107)	(.124)	(.108)	(.118)	(.102)	(.113)	(.112)
(b) Positive Axis Only												
Marsy's Law	-.103	-.0165	-.058	-.0172	-.136	.033	-.00462	-.23*	-.328***	.135	-.351***	-.122
	(.109)	(.114)	(.1)	(.116)	(.119)	(.0938)	(.128)	(.123)	(.111)	(.123)	(.114)	(.111)
(c) Negative Axis Only												
Marsy's Law	-.0755	.0253	-.0552	.0219	.00309	.0101	.0148	-.167	-.288**	-.254**	-.32***	-.204*
	(.128)	(.103)	(.105)	(.102)	(.102)	(.093)	(.118)	(.111)	(.113)	(.111)	(.11)	(.108)
(d) Top 10 Words Only												
Marsy's Law	-.0259	-.0397	-.0101	-.00038	-.0692	.0592	-.194	-.0333	.261**	.33***	-.28**	.489***
	(.118)	(.101)	(.114)	(.102)	(.115)	(.112)	(.121)	(.103)	(.115)	(.1)	(.117)	(.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-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.

A.5 LDA Topic Modelling Evidence

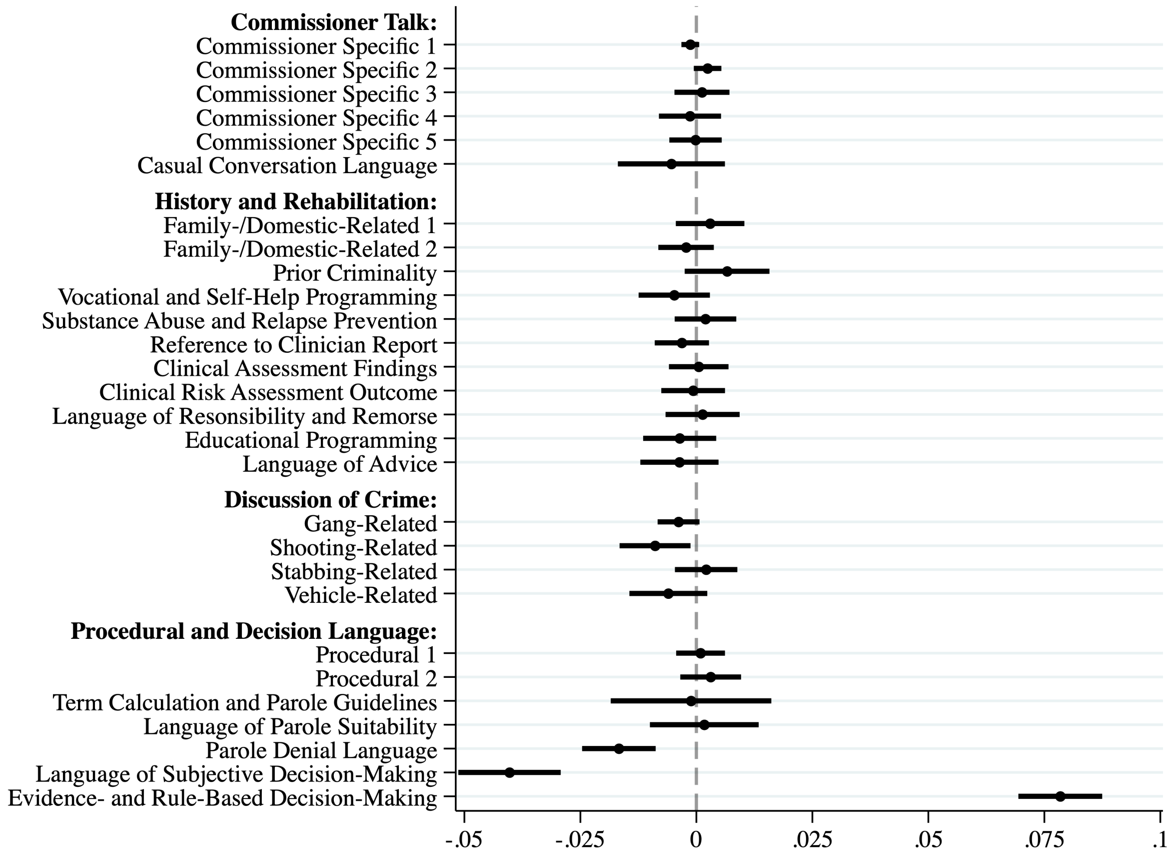
In addition to our semantic axis approach, which takes a deductive approach, we implement Latent Dirichlet Allocation (LDA) topic modelling in order to probe commissioner decision language for additional, novel dimensions of speech that may change in response to Marsy's Law. We train a topic model with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

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 concomittant 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. Commissioners appear to respond to the law by moving away from subjective consideration of the case at hand, to a more objective, rule-based decision-making style. One reading of this combined evidence is that as the denial length-based stakes have increased post-Law, commissioners rely more heavily on objective considerations when making decisions, either to assuage any moral concerns of the harsher denial length decisions, or to insulate themselves from the risk of their decision being appealed or overturned.

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.

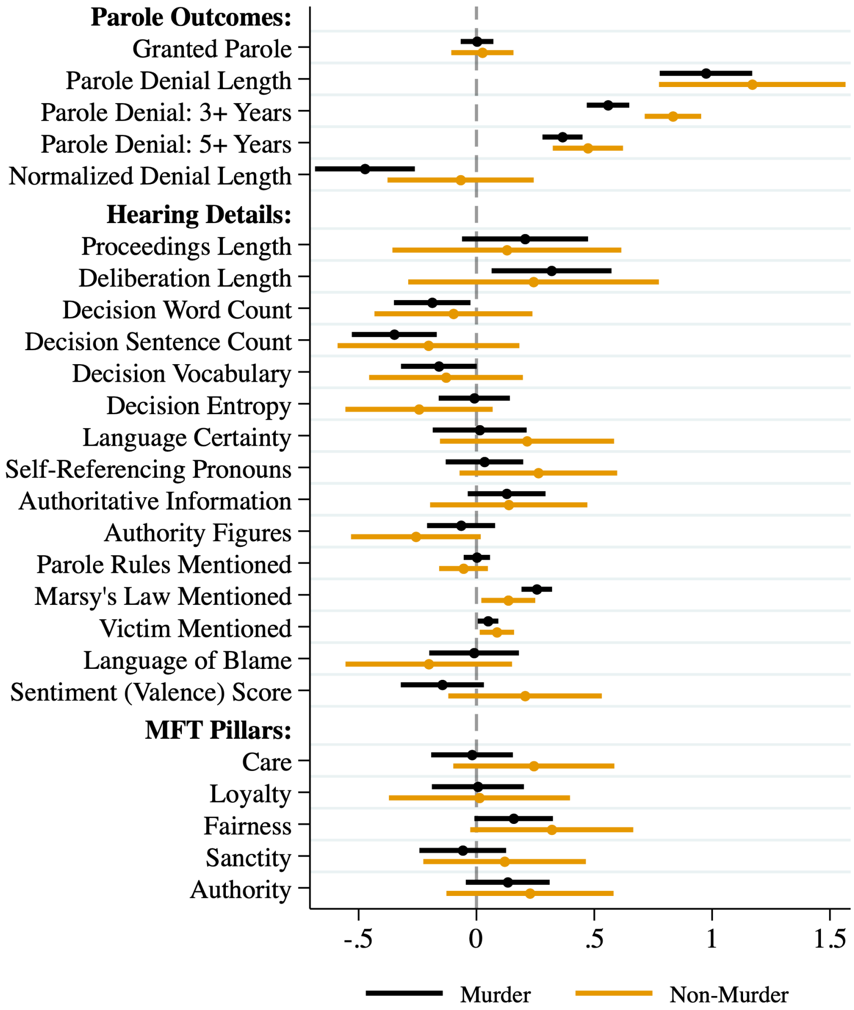
Table A3: Most Frequent Words in LDA Topics

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Comissionner 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

Figure A5: Heterogeneity Analysis: Most Serious Offense



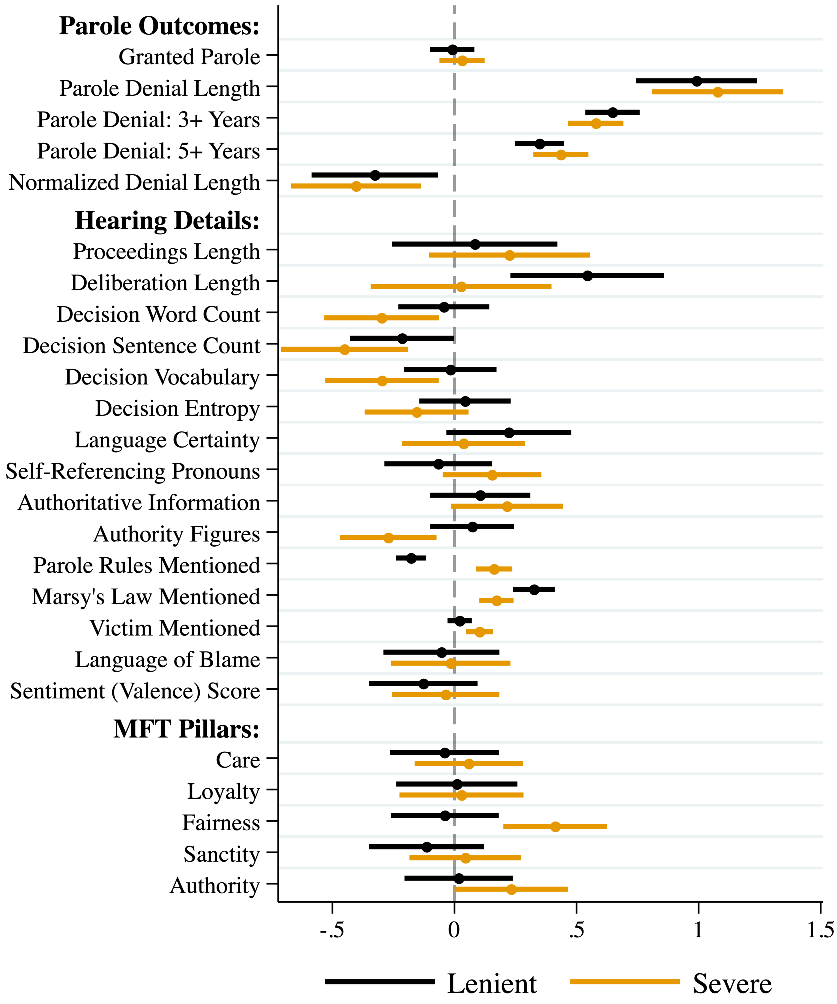
Notes: The coefficient plot presents coefficients and 95% confidence intervals base on Eicker-Huber-White standard errors for the indicator variable Marsy's Law, for both parole outcomes and parole hearings details. All continuous variables are standardized in order to achieve a common scale. Marsy's 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.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 labelled "lenient," their counterparts are labelled 'severe'. We highlight four key findings from this analysis. First, commissioners respond homogeneously to the parole 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

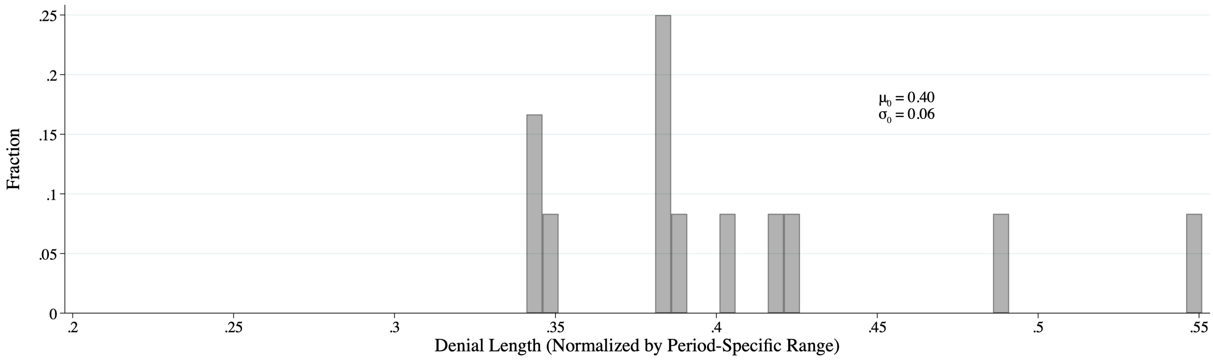


Notes: The coefficient plot presents coefficients and 95% confidence intervals base on Eicker-Huber-White standard errors for the indicator variable Marsy's Law, for both parole outcomes and parole hearings details. All continuous variables are standardized in order to achieve a common scale. Marsy's 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.

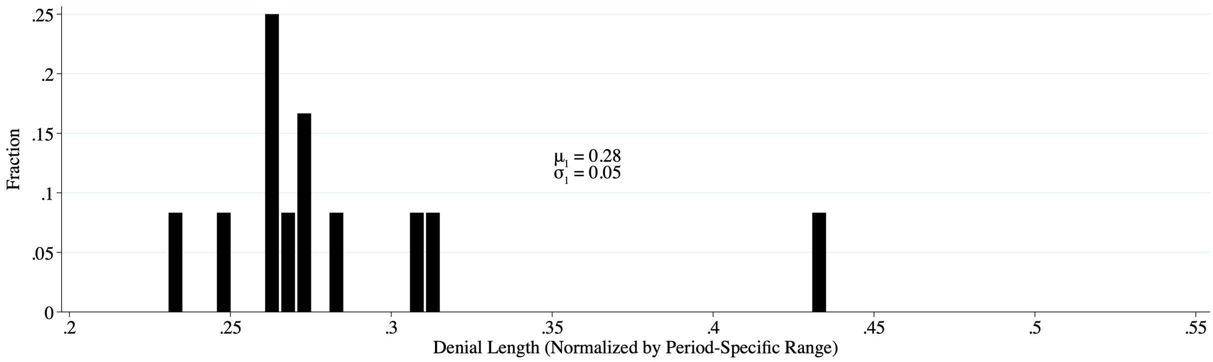
A.7 Inter-Commissioner Variation in Denial Length

As a final piece of analysis, we investigate the extent to which inter-commissioner variation in parole denial length changes in response to Mary'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

Figure A7: Inter-Commissioner Variation in Denial Length



(a) Pre-Marsy's Law



(b) Post-Marsy's Law

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.

this analysis. First, as 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 post-Marsy's Law. 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 compare to prior to the implementation of Marsy's Law. This uniform response to the more onerous denial length options mandated by the Law serves to validate the conceptual approach in this work.