# USING MACHINE LEARNING TO SCRUTINIZE PAROLE RELEASE HEARINGS

*Kristen Bell,*[†] *Jenny Hong,*[††] *Catalin Voss,*[†††] *Graham Todd*[††††] *& AJ Alvero*[†††††]

## ABSTRACT

This Article applies a new approach to machine learning in criminal law: rather than building technological tools to predict the behavior of people put through the system, we build tools to scrutinize the decisions and speaking patterns of those who wield power in the system. We deploy our approach in the context of parole release hearings, using natural language processing to analyze thirty-five thousand hearing transcripts which span approximately seven hundred million total words. We find that when the official presiding over a parole release hearing has a historically lower grant rate, the odds of being granted parole are also lower. We further find that privately retained attorneys speak in more sophisticated legal language than appointed attorneys, and that this more sophisticated language is associated with increased odds of parole. In addition to advancing knowledge about parole release decisions, the Article demonstrates how to leverage machine learning as a tool for uncovering patterns of potential inequity in the criminal legal system.

## TABLE OF CONTENTS



---

DOI: https://doi.org/10.15779/Z38CZ32652
 This work would not have been possible without advising, feedback, and support from Nick McKeown and Johan Ugander. We are thankful for grant support from Stanford's Institute for Human Centered Artificial Intelligence and research assistance from students at the University of Oregon. Catalin Voss was supported by a Google fellowship. We thank Rob Voigt, Chris Potts, and Dan Jurafsky for their assistance with the design of our computational linguistic methods. We thank Chris Manning for his guidance on our NLP information extraction methodology. We thank Terry Winograd for his advice and comments throughout. We thank Julian Nyarko, Jake Searcy, Ryan Sakoda, and the editors of the Berkeley Technology Law Journal for their feedback and editing assistance.

† Assistant Professor at University of Oregon School of Law.
†† Research Scientist at Meta; work on this Article completed while earning PhD at Stanford University.
††† Co-founder of Ello; PhD student at Stanford University (on leave).
†††† Doctoral Candidate; New York University; Tandon School of Engineering.
††††† Assistant Research Professor; Cornell University; Center for Data Science for Enterprise and Society.

## I. INTRODUCTION

Much of the machine-learning technology that is applied to the criminal legal system is focused on predicting the likelihood that an individual will commit violence in the future. Examples of this kind of predictive technology include algorithms that police use in targeting investigations, that judges apply in bail and sentencing decisions, and that parole boards rely on in deciding whom to keep imprisoned.[1] This predictive technology takes as its object of analysis the people who are processed through the legal system, and it generally neglects to scrutinize how officials wield power in the system. In prior work, we proposed an alternative path for machine learning in criminal law that shifts the focus from the people about whom decisions are made to the decision-making itself.[2] Our approach relies on human beings to first make their own discretionary judgments and, only after those judgments have been made, we use machine learning to find patterns in those decisions and mirror them back. The approach is designed to increase transparency in the generally inscrutable webs of discretionary judgments that pervade criminal law.

This Article puts that theoretical approach into practice. We use natural language processing (NLP) to scrutinize the discretionary judgments made at nearly all California parole release hearings over a period of twelve years. We leverage NLP to analyze transcripts from thirty-five thousand hearings; a corpus of five million pages and seven hundred million words. By analyzing such a large sample and scrutinizing language differences in transcripts, we can answer questions about these hearings that have been beyond the reach of prior parole researchers. In particular, this Article considers the following two questions.

First, to what extent does the commissioner assigned to preside over a given hearing explain variation in parole release decisions? In contrast to other states where parole release decisions are made by multiple people, California parole decisions are predominantly made by a single commissioner of the parole board, in consultation with a deputy commissioner. Over fifty different commissioners presided over hearings included in this study. Given the difficulty of proving causal relationships, we cannot conclude that a different

1. Andrew G. Ferguson, *Illuminating Black Data Policing*, 15 OHIO ST. J. CRIM. L. 503, 505 (2018); Lindsey Barrett, *Reasonably Suspicious Algorithms: Predictive Policing at the United States Border*, 41 N.Y.U. REV. L. & SOC. CHANGE 327, 335 (2017); Sharad Goel, Justin M. Rao & Ravi Shroff, *Personalized Risk Assessments in the Criminal Justice System*, 106 AM. ECON. REV.: PAPERS & PROC. 119 (2016); Jon Kleinberg, Himabindu Lakkaraju, Jure Leskovec, Jens Ludwig & Sendhil Mullainathan, *Human Decisions and Machine Predictions*, 133 Q.J. ECON. 237 (2018); State v. Loomis, 881 N.W.2d 749, 755 (Wis. 2016).

2. Kristen Bell, Jenny Hong, Nick McKeown & Catalin Voss, *The Recon Approach: A New Direction for Machine Learning in Criminal Law*, 36 BERKELEY TECH. L.J. 821 (2021).

commissioner would have caused a different outcome in any given case. Our uniquely large sample size and rich set of variables, however, allows us to describe how much variability in decisions is attributable to differences in commissioners rather than differences in underlying case factors. We find that a presiding commissioner with a historically lower grant rate is associated with lower odds that parole will be granted.

Second, we investigate the relationship between retaining a private attorney and a parole candidate's likelihood of being granted parole. California parole candidates have a right to retain an attorney for their parole hearing. If they are indigent, the Board of Parole Hearings ("the Board") appoints and pays for an attorney. We find that representation by a Board-appointed attorney rather than a privately retained attorney is associated with significantly lower odds of parole. Prior studies have been limited in their ability to quantitatively investigate why having a Board-appointed attorney might be associated with a reduced likelihood of parole. Here, we find both that there are significant differences in the words that Board-appointed and privately retained attorneys use at hearings, and that those differences are significant in predicting the outcome of a given hearing.

Our research questions regarding commissioners and attorneys are related in at least two ways. First, the need for extensive resources to compile datasets about parole release decisions has historically limited progress on these research questions. NLP opens new avenues to explore these questions by catalyzing data extraction from unstructured text, allowing for a far larger sample size and a richer set of variables. Furthermore, investigating these two research questions provides insight into arbitrariness across discretionary parole release decisions. One of the longest-standing critiques of parole release decisions is that they are arbitrary.[3] Stated generally, the critique is that, too often, parole candidates with substantially similar case factors do not receive the same outcomes; instead, too often, outcomes are dependent not on case factors but on the idiosyncrasies of the officials who wield power in the parole

3. *See, e.g.*, Fiona Doherty, *Indeterminate Sentencing Returns: The Invention of Supervised Release*, 88 N.Y.U. L. REV. 958, 985, 991–93 (2013); John A. Conley & Sherwood E. Zimmerman, *Decision-Making by a Part-Time Parole Board: An Observational and Empirical Study*, 9 CRIM. JUST. & BEHAV. 396, 396–97 (1982); Anne M. Heinz, John P. Heinz, Stephen J. Senderowitz & Mary Anne Vance, *Sentencing by Parole Board: An Evaluation*, 67 J. CRIM. L. & CRIMINOLOGY 1 (1976); Jon O. Newman, *Parole Release Decision-making and the Sentencing Process*, 84 YALE L.J. 810, 816 n.14, 847 (1975); Kenneth C. Davis, DISCRETIONARY JUSTICE: A PRELIMINARY INQUIRY 58 (1969); Sanford H. Kadish, *Legal Norm and Discretion in the Police and Sentencing Processes*, 75 HARV. L. REV. 904, 912 (1962).

process.[4] As a normative matter, which official presides over the hearing and whether an appointed attorney provides representation should not make a difference to the parole release decision. Both variables are orthogonal to considerations of public safety and outside the control of parole candidates themselves.

Importantly, there are several other variables that normatively should not make a difference in outcomes, such as race, and arguably, the presence of the victim[5] and the district attorney.[6] These other aspects deserve considerable attention, and we address them in future research. In this initial study, the focus is on commissioners and attorneys because the Board itself wields direct power over these variables. Closer examination of these two variables thus provides special insight into how the Board wields its own discretionary power.

This Article proceeds in eight Parts. Part II summarizes our approach to using machine learning as a tool for scrutinizing and improving the exercise of discretion in criminal law. Part III provides background on the parole context in which we deploy our approach and discusses how social science research in the parole context has been limited by the incredibly labor-intensive task of pulling data from unstructured text. Part IV describes the particular laws and policies that govern the California parole system, which our study analyzes. Part V sets forth the method for constructing our dataset, with particular attention to describing how we use NLP to extract data from transcripts. Part VI presents findings showing that NLP is a reliable and effective method to catalyze analysis of parole hearing transcripts. Parts VII and VIII respectively present our substantive findings about the role of parole commissioners and attorney-type at parole hearings. Part IX discusses limitations and future research, and we conclude by returning to reflect more broadly on our approach to applying machine learning in criminal law.

4. *See, e.g.*, Kristen Bell, *A Stone of Hope: Legal and Empirical Analysis of California Juvenile Lifer Parole Decisions*, 54 HARV. C.R.-C.L. L. REV. 455, 460 (2019); W. David Ball, *Heinous, Atrocious, and Cruel: Apprendi, Indeterminate Sentencing, and the Meaning of Punishment*, 109 COLUM. L. REV. 893, 970 n.401 (2009).

5. *See* Edward E. Rhine, Joan Petersilia & Kevin R. Reitz, *The Future of Parole Release*, 46 CRIME & JUST. 279, 318 (2017).

6. *See* R. Michael Cassidy, *Undue Influence: A Prosecutor's Role in Parole Proceedings*, 16 OHIO ST. J. CRIM. L. 293, 302 (2019).

## II. A NOVEL APPROACH TO MACHINE LEARNING IN CRIMINAL LAW

In a prior Article, *The Recon Approach*, we set forth our approach to applying machine learning in the realm of criminal law.[7] The Recon Approach focuses on post-hoc analysis of discretionary legal decisions rather than predicting the behavior of individuals processed through the criminal legal system. Two projects are central to the Recon Approach: reconnaissance and reconsideration. The aim of reconnaissance is to illuminate patterns in how decision-makers tend to make decisions. With increased transparency into decision-making, the public is better positioned to normatively consider the extent to which a system of decision-making may need structural reform. Whereas reconnaissance focuses on patterns across a large set of decisions, reconsideration focuses on individual cases. The objective of reconsideration is to create an ongoing and updated list of decisions that appear to be anomalous within a given set of cases. The decisions on that list would then receive a second-look review by an oversight board and/or appellate attorneys. The second-look review would by no means guarantee a change in the decision, but it would provide an opportunity to reconsider the facts of the case in light of awareness that otherwise like cases tended to receive different outcomes. Working together, we take reconnaissance and reconsideration to provide data-driven opportunities (not guarantees) for improving human discretionary judgment on both a systemic and individual level.

In *The Recon Approach*, we considered various ethical, technological, and political challenges to our approach.[8] A primary ethical concern is that reconsideration tools, like predictive technology, may perpetuate existing inequities. Because reconsideration tools are designed to look for inconsistencies within existing patterns in a set of decisions, the tools will tend to reify whatever patterns are present. Such reification is deeply problematic insofar as the existing patterns in a decision set exhibit inequity. For example, insofar as parole candidates who do not have money to retain private attorneys are generally more likely to be denied parole, then a reconsideration tool would be less likely to flag as anomalous a decision to deny parole to a person who lacks a privately retained attorney. Therefore, conducting reconnaissance to identify existing patterns in decision sets is essential before developing any reconsideration tools. Insofar as reconnaissance uncovers patterns that suggest a risk of inequity, reconsideration tools should be tailored in a way that addresses those risks. This Article, our first in implementing our approach, is

7. Bell et al., *supra* note 2.

8. *Id.*

therefore focused on reconnaissance before we undertake the construction of any reconsideration tools.

Although the reconnaissance aspect of our approach is novel in its use of NLP,[9] it builds on a well-established body of quantitative social science that is focused on uncovering patterns in legal decision-making.[10] As discussed in the next Part, large-scale quantitative studies have been limited in the parole context due to the incredibly labor-intensive task of pulling data from unstructured text. NLP offers a reliable and effective tool to catalyze research in this area.

## III. BACKGROUND ON PAROLE AND PAROLE RESEARCH

Approximately half a million people are released on parole per year across the United States,[11] with discretionary decisions by parole boards accounting for an estimated one-third to one-half of all releases.[12] The laws, policies, and general functioning of parole boards vary considerably across states.[13] One feature that is shared across nearly all parole systems is that parole boards exercise remarkably broad discretion. The U.S. Supreme Court has characterized parole decisions as a "necessarily subjective" inquiry; a "discretionary assessment of a multiplicity of imponderables, entailing primarily what a man is and what he may become rather than simply what he has done."[14] In the words of renowned criminal law scholar Sanford Kadish, officials who make parole decisions have "the greatest degree of uncontrolled power over the liberty of human beings that one can find in the legal system."[15] The legal standards governing parole are remarkably vague; many state statutes

---

9. As of 2024, we are aware of two other studies that apply NLP to scrutinize patterns in decision-making in the context of criminal law. *See* Anna Effenberger, John H. Blume & Martin T. Wells, *Quantifying Disparate Questioning of Black and White Jurors in Capital Jury Selection*, 20 J. EMPIRICAL LEGAL STUD. 609 (2023) (applying NLP to voir dire questioning in death penalty cases); Hannah Laqueur & Anna Venancio, *Computational Analysis of California Parole Suitability Hearings*, in LAW AS DATA: COMPUTATION, TEXT, & THE FUTURE OF LEGAL ANALYSIS 193, 207–08 (Michael A. Livermore & Daniel N. Rockmore, eds., 2019).

10. *See, e.g.*, David C. Baldus, George Woodworth & Charles A. Pulaski Jr., EQUAL JUSTICE AND THE DEATH PENALTY: A LEGAL AND EMPIRICAL ANALYSIS (1990); Andrew Gelman, Jeffrey Fagan & Alex Kiss, *An Analysis of the New York City Police Department's "Stop-and-Frisk" Policy in the Context of Claims of Racial Bias*, 102 J. AM. STAT. ASS'N 813 (2007); David Arnold, Will Dobbie & Crystal S. Yang, *Racial Bias in Bail Decisions*, 133 Q.J. ECON. 1885 (2018).

11. Barbara Oudekerk & Danielle Kaeble, U.S. DEP'T OF JUST., NCJ 256092, PROBATION & PAROLE IN THE U.S., 2019, at 9 (2021).

12. Kimberly A. Thomas & Paul D. Reingold, *From Grace to Grids: Rethinking Due Process Protections for Parole*, 107 CRIM. L. & CRIMINOLOGY 213, 239 (2017).

13. Rhine et al., *supra* note 5, at 279.

14. Greenholtz v. Inmates of Neb. Penal & Corr. Complex, 442 U.S. 1, 11 (1979).

15. Kadish, *supra* note 3, at 916.

allow parole boards to grant parole if doing so serves the "best interest of society" or if there is a "likelihood" that the person will not violate the law upon release.[16] Given the open-ended, discretionary nature of parole decisions, critics have argued that parole boards fail to treat like cases alike,[17] that decisions exhibit racial disparities,[18] and that grant rates shift with the political tide.[19]

Some statistical evidence supports these critiques,[20] but large-scale computational studies of parole are relatively rare compared to other parts of the American criminal legal system, such as policing,[21] pre-trial detention,[22] and sentencing.[23] Quantitative studies of parole release decisions generally involve sample sizes of fewer than one thousand cases,[24] although there are a few notable exceptions.[25] The studies that are most relevant to our project have focused on California parole release decisions. One prior study used NLP to extract fourteen variables from over eight thousand California parole hearing

---

16. *See* Kristen Bell, *The Forgotten Jurisprudence of Parole and State Constitutional Doctrines of Vagueness*, 44 CARDOZO L. REV. 1953, 2004 (2023).

17. *See* Bell, *supra* note 4, at 460.

18. *See* Kathryne M. Young & Jessica Pearlman, *Racial Disparities in Lifer Parole Outcomes: The Hidden Role of Professional Evaluations*, 47 L. & SOC. INQUIRY 783 (2022); Michael Winerip, Michael Schwirtz & Robert Gebeloff, *For Blacks Facing Parole in New York State, Signs of a Broken System*, N.Y. TIMES (Dec. 4, 2016), https://www.nytimes.com/2016/12/04/nyregion/new-york-prisons-inmates-parole-race.html; Beth M. Huebner & Timothy S. Bynum, *The Role of Race and Ethnicity in Parole Decisions*, 46 CRIMINOLOGY 907 (2008).

19. *See, e.g.*, Kevin R. Reitz & Edward E. Rhine, *Parole Release and Supervision: Critical Drivers of American Prison Policy*, 3 ANN. REV. CRIMINOLOGY 281, 286 (2020).

20. *See supra* notes 17–19.

21. Gelman et al., *supra* note 10; *see* Emma Pierson, Camelia Simoiu, Jan Overgoor, Sam Corbett-Davies, Daniel Jenson, Amy Shoemaker, Vignesh Ramachandran, Phoebe Barghouty, Cheryl Phillips, Ravi Shroff & Sharad Goel, *A Large-Scale Analysis of Racial Disparities in Police Stops Across the United States*, 4 NATURE OF HUM. BEHAV. 736 (2020).

22. Arnold et al., *supra* note 10.

23. *See* David S. Abrams, Marianne Bertrand & Sendhil Mullainathan, *Do Judges Vary in Their Treatment of Race?*, 41 J. LEGAL STUD. 347 (2012); James M. Anderson, Jeffrey R. Kling & Kate Stith, *Measuring Interjudge Sentencing Disparity: Before and After the Federal Sentencing Guidelines*, 42 J.L. ECON. 271 (1999); Stephen Klein, Joan Petersilia & Susan Turner, *Race and Imprisonment Decisions in California*, 247 SCI. 812 (1990).

24.*See* Joel M. Caplan, *What Factors Affect Parole: A Review of Empirical Research*, 71 FED. PROB. 16, 16 (2007).

25. *See* Shamena Anwar & Hanming Fang, *Testing for Racial Prejudice in the Parole Board Release Process: Theory and Evidence*, 44 J. LEGAL STUD. 1, 23 (2015); Stéphane Mechoulan & Nicolas Sahuguet, *Assessing Racial Disparities in Parole Release*, 44 J. LEGAL STUD. 39, 61 (2015); Gerald G. Gaes & Julia Laskorunsky, ROBINA INST. OF CRIM. L. & CRIM. JUST., FACTORS AFFECTING COLORADO PAROLE RELEASE DECISIONS, 8 (2022); Hannah Laqueur & Ryan Copus, *An Algorithmic Assessment of Parole Decisions*, 40 J. QUANTITATIVE CRIMINOLOGY 151, 162 (2022); Laqueur & Venancio, *supra* note 9, at 195.

transcripts in the time period of 2011 to 2014.[26] The study found that the risk assessment score is the strongest predictor of whether a candidate will be found suitable for parole.[27] The study also found that the presence of an interpreter at the hearing, the presence of victims at the hearing, and a conviction for a sex crime are strongly associated with decisions to deny parole.[28] The time of day of the hearing was not significant in predicting hearing outcomes.[29] Through text analysis, the study found that candidates granted parole were more likely to use polite and deferential language.[30]

Other studies required researchers to read a small subset of hearing transcripts and meticulously code for variables of interest.[31] The scope of these prior studies ranged from a total of 107 to 754 transcripts and analyzed between fourteen and twenty-one variables. The first of these studies considered a sample of 754 transcripts from lifer parole release hearings held from 2007 to 2010.[32] The California Department of Corrections and Rehabilitation (CDCR) provided some data about these hearings, but the majority of data were pulled from transcripts by research assistants. A logistic regression analysis considered twenty-one independent variables.[33] The study found that the following seven factors were associated with an increased likelihood of being granted parole: younger age at the time of the crime, older age at the time of the hearing, a "low" risk score on a forensic psychology evaluation, fewer disciplinary infractions in prison, a confirmed job offer, participation in a substance abuse program, and a history of violent crime before the commitment offense.[34] The following three factors were found to be associated with a decreased likelihood of being granted parole: failure to adequately answer a question about the Twelve Steps of Alcoholics/Narcotics Anonymous, opposition from the district attorney, and attempts to evade law enforcement officers after commission of the offense.[35]

26. *See* Laqueur & Venancio, *supra* note 9, at 193, 205–06, 221.
27. *Id.* at 207.
28. *Id.* at 210–12.
29. *Id.* at 219.
30. *Id.* at 223, 228.
31. *See* Bell, *supra* note 3, at 455; Beth Caldwell, *Creating Meaningful Opportunities for Release: Graham, Miller and California's Youth Offender Parole Hearings*, 40 N.Y.U. REV. L. & SOC. CHANGE 245, 245 (2016); David R. Friedman & Jackie M. Robinson, *Rebutting the Presumption: An Empirical Analysis of Parole Deferrals Under Marsy's Law*, 66 STAN. L. REV. 173 (2014); Kathryne M. Young, Debbie A. Mukamal & Thomas Favre-Bulle, *Predicting Parole Grants: An Analysis of Suitability Hearings for California's Lifer Inmates*, 28 FED. SENT'G REP. 268 (2016).
32. *See* Young et al., *supra* note 32.
33. *See id.* at 273.
34. *See id.*
35. *See id.*

Another study considered a sample of 680 parole hearings held in the same time period and also relied on data from both CDCR and variables pulled from transcripts by research assistants.[36] Regression analyses in that study considered nineteen independent variables and focused on the role of race and professional assessments. It found that Black candidates had a lower likelihood of parole relative to White candidates.[37] The difference between Black and White candidates' likelihood of parole became statistically insignificant, however, when the regression included variables that measure professional assessments (a low-risk score on a psychological evaluation, no opposition from the district attorney, and fewer disciplinary citations).[38] The study noted that extensive time and resources were needed to pull information from transcripts and that a larger sample size would have allowed for deeper investigation.[39]

A third study of California parole hearings investigated a sample of 302 decisions to deny parole during 2011.[40] CDCR provided some data about all these 302 decisions, and researchers coded transcripts from a sub-sample of 103 decisions to gain more information. Using regression analysis, the study found the following variables were significant in predicting the deferral period (how many years a denied parole candidate would have to wait until their next parole hearing): the parole candidate's gender, representation by a privately retained attorney, the commissioner's identity, mental health history, prison security level, and psychological risk assessment score.[41]

Two studies have considered California parole hearings for parole candidates who were under eighteen years old at the time of the offense. One study considered 107 hearings in 2014,[42] and the other considered 465 hearings in 2014–2015.[43] The larger study focused on assessing the degree of consistency across cases with respect to an operationalized measure of rehabilitation.[44] It found that a significant degree of variability across decisions was attributable to variables that are not relevant to rehabilitation, including

36. *See* Young & Pearlman, *supra* note 18, at 794–95.
37. *See id.* at 805–07.
38. *See id.*
39. *See id.* at 812.
40. *See* Friedman & Robinson, *supra* note 31, at 190.
41. *See id.* at 195, 203.
42. *See* Caldwell, *supra* note 31.
43. *See* Bell, *supra* note 5, at 473.
44. *See id.* at 480 (the measure of rehabilitation was based on "post-conviction behavior over which parole candidates have reasonable control—for example, participation in programs that are offered at the prison and a pattern of compliance with prison rules").

race, attorney-type, victim and district attorney opposition, and prior experience appearing before the Board.[45]

## A. ROLE OF PAROLE BOARD MEMBERS AS ADJUDICATORS

Researchers have sought to measure the extent to which variability in court decision-outcomes is attributable to differences in decision-makers. Several studies have found significant inter-judge disparities in the context of criminal sentencing decisions.[46] Furthermore, researchers have found significant disparities across adjudicators in social security decisions and immigration cases.[47] Such findings regarding inter-adjudicator variability have been important considerations in adjusting legal policy to improve consistency in decision-making.[48]

Research on inter-adjudicator disparities in the parole context, however, has been relatively limited. Parole release hearings are well-suited for this type of analysis, particularly given the broad discretion that parole boards currently exercise and the increasing calls for reform to increase consistency in outcomes across cases.[49] Some scholars point to large fluctuations in the overall rates of parole release as evidence that parole release decisions are influenced more on "personal ideologies of parole authority members, and those who appoint them, than from objective factors."[50] Qualitative research has explored parole board members' thought processes in decision-making through interviews, surveys, and the rhetoric in their reasons for decisions.[51] These studies identify

45. *See id.* at 460.

46. *See, e.g.*, Crystal S. Yang, *Have Interjudge Sentencing Disparities Increased in an Advisory Guidelines Regime? Evidence from Booker*, 89 N.Y.U. L. REV. 1268 (2014); Joshua Fischman & Max M. Schanzenbach, *Do Standards of Review Matter? The Case of Federal Criminal Sentencing*, 40 J. LEGAL STUD. 405–37 (2011); Abrams et al., *supra* note 23.

47. *See* Jerry L. Mashaw, Charles J. Goetz, Frank I. Goodman, Warren F. Schwartz, Paul R. Verkuil & Milton M. Carrow, SOCIAL SECURITY HEARINGS AND APPEALS: A STUDY OF THE SOCIAL SECURITY ADMINISTRATION HEARING SYSTEM (1978); Jaya Ramji-Nogales, Andrew I. Schoenholtz & Philip G. Schrag, *Refugee Roulette: Disparities in Asylum Adjudication*, 60 STAN. L. REV. 295 (2007); Joshua B. Fischman, *Measuring Inconsistency, Indeterminacy, and Error in Adjudication*, 16 AM. L. & ECON. REV. 40 (2014).

48. *See* Fischman, *supra* note 47.

49. *See* Kevin R. Reitz, Edward E. Rhine, Allegra Lukac & Melanie Griffith, ROBINA INST. OF CRIM. L. & CRIM. JUST., AMERICAN PRISON-RELEASE SYSTEMS: INDETERMINACY IN SENTENCING AND THE CONTROL OF PRISON POPULATION SIZE (2022).

50. Mario A. Paparozzi & Roger Guy, *The Giant That Never Woke: Parole Authorities as the Lynchpin to Evidence-Based Practices and Prisoner Reentry*, 25 J. CONTEMP. CRIM. JUST. 397, 404 (2009).

51. *See* David P. Connor, *How to Get Out of Prison: Views from Parole Board Members*, 1 CORR. 107 (2016); Kathryne M. Young & Hannah Chimowitz, *How Parole Boards Judge Remorse: Relational Legal Consciousness and the Reproduction of Carceral Logic*, 56 LAW & SOC'Y REV. 237

common themes and decision-making styles that tend to be shared among parole board members, rather than attempt to quantify degrees of difference across individual board members.

Quantitative research on inter-adjudicator disparities in the parole context is limited to two studies. Laqueur and Venancio found that "at least 11% of cases would be decided differently based on the presiding commissioner that happens to be assigned to an inmate's case, and at least 15% could be decided differently depending only on the deputy commissioner assigned to hear the case."[52] Additionally, Friedman and Robinson found some evidence that the identity of presiding parole commissioners was significant in predicting the length of deferral periods, but their evidence was not conclusive.[53]

### B. Role of Attorneys who Represent Parole Candidates

The U.S. Constitution does not guarantee the right to be represented by an attorney in the context of a parole release decision. Some states allow attorneys to represent people in parole proceedings, while other states prohibit the presence of attorneys at parole hearings.[54] California recognizes not only a right to have an attorney at a parole hearing, but also a right to an appointed attorney in the event that a parole candidate cannot retain a private attorney.[55]

Three studies have analyzed the role of attorneys in parole release decisions. The first study, by Friedman and Robinson, found that having a Board-appointed attorney was associated with an increased likelihood of a longer deferral period at California parole hearings.[56] The second study, by Bell, found that having a Board-appointed attorney was associated with a reduced likelihood of parole at California parole hearings for juvenile lifers.[57] The third study, by Kokkalera, found different results at juvenile lifer parole hearings in a state other than California. In that other state, having an appointed attorney as compared to a retained attorney was not significant in

(2022); Jon'a F. Meyer, *Strange Science: Subjective Criteria in Parole Decisions*, 24 J. CRIME & JUST. 43 (2001); Joss Greene & Isaac Dalke, *"You're Still an Angry Man": Parole Boards and Logics of Criminalized Masculinity*, 25 THEORETICAL CRIMINOLOGY 639 (2021).

52. Laqueur & Venancio, *supra* note 9, at 216–17.

53. *See* Friedman & Robinson, *supra* note 31, at 197.

54. *See* Edward E. Rhine, Joan Petersilia & Kevin R. Reitz, *Improving Parole Release in America*, 28 FED. SENT'G REP. 96, 100 (2015).

55. *See* CAL. PENAL CODE § 3041.7 (West 2024); CAL. CODE REGS. tit. 15, § 2256 (c) (2023).

56. *See* Friedman & Robinson, *supra* note 31, at 197.

57. *See* Bell, *supra* note 4 at 500.

predicting the parole decision.[58] Further, having an appointed attorney was associated with a reduced likelihood of a longer deferral period.[59] The authors of these respective studies noted limitations in their analysis regarding attorney-type, and raised concerns that findings about attorney-type may be attributable to certain types of selection bias.[60]

Despite the limitations, the different findings in these respective studies are nevertheless puzzling; why does having an appointed attorney appear to disadvantage parole candidates in California whereas it appears to advantage parole candidates in the state studied by Kokkalera? Part of the answer could be due to differences in the parole systems across states. In addition, Kokkalera suggested that the difference in the findings could be attributable to differences in how appointment of counsel is structured and regulated. In the state studied by Kokkalera, the state's public defender agency appoints and regulates parole attorneys. That agency has detailed guidelines for attorneys in parole cases and does not impose a fixed cap on fees, expenses, or time spent on parole cases. In contrast, a public defense agency does not govern appointed attorneys for California parole hearings. Instead, parole attorneys in California are appointed and compensated by the Board; during the time period of the cited studies, these attorneys were paid a flat rate of $400 per case.[61] Both training and compensation rates have since increased,[62] following a lawsuit which argued that Board-appointed attorneys provide ineffective assistance of counsel.[63]

## IV. STUDY SETTING, CALIFORNIA LIFER PAROLE HEARINGS

In California, most individuals who are eligible for discretionary parole release hearings are serving life with the possibility of parole sentences

58. *See* Stuti S. Kokkalera, *Representing Juvenile Lifers: Do Attorneys in Parole Hearings Matter?*, 45 J. CRIME & JUST. 189 (2021). The article does not provide the name of the state, explaining that "[d]ue to ongoing projects with the state parole board, the state has been de-identified." *Id.* at 14 n.1.

59. *See id.*

60. *E.g.*, Bell suggested that attorney-type likely correlates with relatively lower socio-economic status. *See* Bell, *supra* note 4, at 488. Kokkalera noted that more than half of the retained attorneys in the study sample were student attorneys. *See* Kokkalera, *supra* note 58, at 200.

61. *See* Bell, *supra* note 4, at 509.

62. Bd. of Parole Hearings, 2023 Report of Significant Events, 11–12 (2024) (on file with authors).

63. *See In re* Poole, No. A154517, 2018 WL 3526684 (Cal. Ct. App. July 23, 2018).

(“lifers”).[64] California incarcerates more than thirty-three thousand lifers, which is about one-third of all lifers in the United States.[65] Lifers make up approximately 27% of California’s prison population.[66] California is an ideal site for this study due to the large population of individuals eligible for parole and because of the fact that each hearing produces a transcript.

California law provides that the Board “shall normally” grant parole after a parole candidate has served the minimum period of incarceration period required by the sentence,[67] unless the Board determines that the candidate “continues to pose an unreasonable risk to public safety.”[68] The Board follows administrative regulations that outline, among other things, factors that generally support a candidate’s suitability or unsuitability for parole.[69] While the administrative regulations provide guidance, the ultimate question is “current dangerousness.”[70] If the Board finds that the parole candidate does not pose a current danger to the community, parole must be granted.[71] The facts of the crime and social history prior to the crime cannot, on their own, support a denial of parole.[72] Such facts can, however, support a denial of parole if there is a “rational nexus” between the crime and current attitudes or recent conduct.[73]

The following procedural rights are provided at California parole release hearings: a notice of the hearing, a review of the prison file prior to the hearing,[74] legal counsel,[75] and appointment of legal counsel if the parole candidate is indigent.[76] The Board itself appoints and pays counsel, in contrast to criminal proceedings in which courts appoint public defenders. There is no right to a hearing in public; media and members of the public may observe

64. *See* Cal. Bd. of Parole Hearings, *Discretionary Parole in California: Report for the Committee on Revision of the Penal Code*, 9–11 (2020), https://www.cdcr.ca.gov/bph/wp-content/uploads/sites/161/2023/05/pv-Discretionary-Parole-in-California-November-2020.pdf (For decades prior to 2014, the only individuals eligible for discretionary release on parole were serving life with the possibility of parole sentences; legal changes over the past ten years have expanded eligibility to other people serving long sentences.).

65. *See* Ashley Nellis, SENT’G PROJECT, NO END IN SIGHT: AMERICA’S ENDURING RELIANCE ON LIFE IMPRISONMENT, 10 (2021).

66. *See supra* note 62, at 2.

67. CAL. PENAL CODE § 3041(a)(2) (West 2024).

68. *In re Lawrence*, 44 Cal. 4th 1181, 1212, 1221 (2008).

69. *See* CAL. CODE REGS. tit. 15, § 2402 (2023).

70. *See Lawrence*, 44 Cal. 4th at 1210.

71. *See id.* at 1226–27.

72. *See id.*

73. *See id.* at 1227.

74. *See* CAL. PENAL CODE § 3041.5 (West 2024).

75. *See id.* § 3041.7.

76. CAL. CODE REGS. tit. 15, § 2256(c) (2023).

only if the request is approved by the Board.[77] Victims and victims' next-of-kin have a right to be notified about and attend hearings, but friends, family, or other supporters of the parole candidate have no right to attend hearings.[78] Both the public and the parole candidate have a right to transcripts of hearings.[79] State law requires that the transcripts include everything that is said in the hearing and a definitive, exhaustive statement of the reasons for the parole decision.[80]

### A. RECORD OF EVIDENCE AT HEARINGS

The Board considers all relevant and reliable information available in determining parole suitability.[81] Information includes, but is not limited to: records from the underlying conviction; records of misconduct in prison; records of participation in education, vocation, and self-help groups in prison; any essays or self-help book reports that a parole candidate has written; transcripts from prior parole hearings; mental health records; written statements by the candidate; letters of support from family, friends, and community members; written statements of commendation by prison staff ("laudatory chronos"); documentation of parole plans; letters of opposition; and statements by the victim or the victim's next-of-kin.[82]

The Board also considers a "Comprehensive Risk Assessment" (CRA) report. The report is written by a forensic psychologist employed by the Board who conducts an interview with the parole candidate and reviews their prison file shortly before the parole hearing.[83] The psychologist uses a risk-assessment tool, and using their professional judgment, gives the parole candidate a score of low risk, moderate risk, or high risk.[84]

77. *See id.* §§ 2029.1, 2030.

78. *See* CAL. PENAL CODE § 3043; *see* CAL. CODE REGS. tit. 15, § 2029.1 (stating that attendance is only permitted for educational or informational purposes, which would not include family or friends).

79. CAL. PENAL CODE § 3041.5(a)(4); *see In re Bode*, 74 Cal. App. 4th 1002, 1003 (1999).

80, *See In re Prather*, 50 Cal. 4th 238 (2010).

81. *See* CAL. CODE REGS. tit. 15, § 2402(b).

82. *See* generally Cal. Bd. of Parole Hearings, Parole Consideration Transcripts (2007–2019) (on file with authors) (containing records of parole hearings that consider these types of information).

83. *See* CAL. CODE REGS. tit. 15, § 2240 (2015); CAL. BD. OF PAROLE HEARINGS, *Forensic Assessment Division*, https://perma.cc/2GNW-ST2T (last visited Feb. 19, 2025).

84. HEATHER MACKAY & PRISON L. OFF., THE CALIFORNIA PRISON AND PAROLE LAW HANDBOOK 284 (2019), https://prisonlaw.com/wp-content/uploads/2019/01/Handbook-Chapter-9.pdf.

## B. PROCEEDINGS AT PAROLE HEARINGS

Parole hearings are conducted by Board commissioners whom the Governor appoints for three-year terms.[85] Generally, one commissioner, who is a member of the Board (the presiding commissioner), and one deputy commissioner are present to conduct the hearing and decide whether to grant parole.[86] Others present at the hearing generally include the parole candidate, the candidate's attorney, sometimes the district attorney from the county of the underlying conviction, and sometimes victims or victims' next-of-kin.[87]

For most of the hearing, the presiding commissioner and deputy commissioner question the parole candidate. Questions generally fall into four categories: (i) the candidate's background prior to the conviction; (ii) the underlying offense; (iii) post-conviction activities; and (iv) parole plans.[88] Once the questioning has concluded, the district attorney and the parole-candidate's attorney are given the opportunity to ask clarifying questions and make closing statements. A closing statement can then be given by the parole candidate, followed by a statement from the victim or the victim's next of kin.[89]

The presiding commissioner and deputy commissioner deliberate alone, and those deliberations are not included in the transcripts. Then, they invite the parties to return to the hearing room and the transcript begins recording again. The presiding commissioner then announces the decision as to whether the parole candidate is suitable for parole and the reasoning for the decision.[90] If parole is denied, the commissioner also announces how many years the candidate will wait until the next hearing is scheduled (the "deferral period").[91] California law sets the presumptive deferral period at fifteen years, but the Board may set a shorter time period based on considerations of public safety.[92]

During the time period of our study, parole hearings ranged in length from one to three hours and were generally conducted in a room inside the prison

85. CAL. PENAL CODE § 5075(b)(1).

86. *See id.* § 3041(a)(2).

87. *See id.* §§ 3041.7, 3043.

88. *See generally* Cal. Bd. of Parole Hearings, Parole Consideration Transcripts (2007–2019) (on file with authors).

89. *See id.*

90. The Board's decision as to whether a candidate is suitable for parole is not a final decision. After the panel present at the hearing makes a decision, the Board's internal Decision Review Unit reviews decisions and may recommend a modification. CAL. CODE REGS. tit. 15, § 2041(h) (2023). Parole decisions are then referred to the Governor who has the authority to reverse the decision in murder cases. CAL. CONST. art. V, § 8. For simplicity, this Article uses the terms "grant" and "deny" rather than "finds suitable" and "finds unsuitable."

91. *See* CAL. PENAL CODE § 3041.5.

92. *See id.*

where the parole candidate is incarcerated.[93] After the time period of our study, there have been significant changes to proceedings. For example, since 2021, hearings may occur via video conference.[94] Additionally, the Board adopted a structured decision-making framework which may reduce the average time of hearings.[95]

## V. METHOD FOR CONSTRUCTING DATASET

Through a California Public Records Act (CPRA) request initiated in 2018, we obtained and organized a complete corpus of every digitally available parole hearing transcript for candidates in California. The resulting corpus contains 34,993 transcripts and constitutes a complete record of all disclosed hearings from January 2, 2007 to November 22, 2019.[96] The transcripts average 18,499 words each.

In addition to requesting transcripts, we also requested structured data about each hearing, focusing on features that were not reliably included in the transcripts themselves, such as the race/ethnicity of the parole candidate. The CDCR Data Requests team provided limited data on the commitment offense and conviction year for a subset of hearings, but it did not provide the majority of data that we requested. In August 2020, we obtained a court order through the San Francisco Superior Court of California for the release of three features: (1) the race/ethnicity of each parole candidate; (2) the current status of each parole candidate (released, returned to CDCR, or deceased); and (3) the status of the attorney representing the parole candidate at each hearing (whether the attorney was retained by the parole candidate or appointed by the Board).

93. *See* generally Cal. Bd. of Parole Hearings, Parole Consideration Transcripts (2007–2019) (on file with authors) (stating location of hearings as prisons across California with few exceptions).

94. *See* CAL. PENAL CODE § 3041.6.

95. *See* Ralph C. Serin, Kaitlyn Wardrop, Laura Gamwell & Jennifer Shaffer, *Parole Decision Making: Moving Towards Evidence-Based Practice*, *in* HANDBOOK ON MOVING CORRECTIONS AND SENTENCING FORWARD: BUILDING ON THE RECORD 148–49 (Pamela K. Lattimore, Beth M. Huebner & Fa Taxman eds., 2021).

96. We received from CDCR a total of 35,105 transcripts. CDCR indicated that they withheld a small number of transcripts, citing confidentiality concerns. We excluded transcripts in which the result of the hearing could not be extracted, resulting in our total study sample being 34,993 transcripts. The time period begins with 2007 because the Board informed us that hearings prior to 2007 were not regularly digitized into PDF files. The time period ends with Nov. 22, 2019, because this was the date of the most recent transcript that was available when we received the data.

CDCR complied with this court order, providing these three features for all hearings.[97]

Our data is at the hearing level; the total number of parole candidates included in our dataset is 15,852 individuals. The number of individuals is significantly fewer than the total number of hearings because most individuals had multiple hearings.[98] For each hearing, we aimed to extract many pieces of information ("features") which were hypothesized to influence parole decisions. We were as expansive as possible in selecting features; we included all features identified as more than marginally predictive in prior studies of California parole hearings. We also included features suggested in discussions with legal experts in parole, formerly incarcerated individuals, advocacy groups, representatives from the California Governor's office, and the Board.[99] Our list of features included, for example, the rehabilitation programs that parole candidates participated in, their history of disciplinary write-ups in prison, their commitment offense, their risk assessment score, and whether they had a job offer available if released. After creating an initial list of 118 features of interest, we attempted to extract them from transcripts using both manual and automated methods, each of which is described below.

### A. MANUAL EXTRACTION METHOD

With approval from the University of Oregon and the Stanford University Institutional Review Boards (IRBs), a team of ten paid research assistants was recruited to manually annotate a subset of transcripts. All research assistants received IRB training followed by a three-week training process during which they familiarized themselves with a detailed coding manual and a custom-built data annotation tool. The tool included a menu of annotation tasks. For each annotation task, the research assistants extracted the value of the field and clicked on one or more sentences in the hearing transcript from which they identified the information.

During a training round of annotations, the research assistants coded three transcripts for our initial list of 118 features that were hypothesized to influence parole decisions. We then assigned new sets of transcripts to the

97. *See* Verified Petition for Writ of Mandate Ordering Compliance with the California Public Records Act, Voss v. Cal. Dep't of Corrections and Rehabilitation, No. CPF-20-517117 (Cal. Super. Ct. 2020), https://www.eff.org/document/petition-voss-v-cdcr.

98. We use hearings rather than individuals as our unit of analysis because of our interest in investigating the role of the particular commissioner and attorney type. An individual who has multiple hearings is generally assigned a different presiding commissioner for each hearing and may have different attorneys for each hearing.

99. Discussions with the Board include two conversations with Director Jennifer Shaffer in late 2018 and early 2019.

research assistants. Each transcript coded in this round was annotated by three different research assistants. For a subset of the triple-coded transcripts, a legal expert who was part of the study team resolved any conflicts during training sessions. After the first month of annotations, two research assistants were identified as the most reliable annotators. Subsequent transcripts were double-coded by one of these two research assistants plus an additional research assistant. We measured inter-rater reliability using several measures, including Gwet's $AC_1$.[100] We dropped several features due to inadequate inter-rater reliability and/or class imbalance.

## B. AUTOMATED EXTRACTION METHODS

### 1. *Direct Extraction from Title and Closing Pages*

Several features are specified in a relatively structured manner on a title or closing page of the transcript. Those features include the date of the hearing, the prison where the hearing is held, the name and CDCR number of the parole candidate, the name of their attorney,[101] the names of the presiding and deputy commissioners, a list of others in attendance (including, if applicable, the district attorney and/or the victim or victim's next of kin[102]), and the decision outcome. We extracted these features with perfect accuracy using a series of custom PDF parsing tools implemented in Python. To correct for misspelling in the names of commissioners, attorneys, and prisons, we performed clustering using string similarity metrics.

### 2. *Weakly-Supervised Labeling Functions*

We used weakly supervised labeling functions to extract the following features: participation in rehabilitation programs focused on gang activity

100. In studies where each data point is labeled by more than one annotator, it is standard to report the rate at which annotators agree with each other. Gwet's $AC_1$ compares the observed agreement between annotators. If the same data point is always labeled the same, then there is perfect agreement between annotators. Gwet's $AC_1$ corrects for the expected agreement between annotators under chance, much like the idiom that a broken clock is right twice a day. To be more precise, chance agreement refers to the rate of agreement that occurs if two annotators are independently randomly guessing at labels. *See* Kilem Li Gwet, HANDBOOK OF INTER-RATER RELIABILITY 101–28 (4th ed. 2010).

101. Although the name of the attorney is listed, the name alone is insufficient for determining whether the attorney is Board-appointed or privately retained, because the same attorney may serve in either the Board-appointed or the privately retained role during the timespan of our transcripts.

102. In addition to extracting whether the district attorney and/or victim or victim's next of kin were present, we tried to extract whether these individuals opposed parole. We tried to extract opposition using sentiment analysis, but it failed to outperform the measure of presence. This is likely due to the strong class imbalance; victim representatives and district attorneys almost always oppose parole if they attend a hearing.

("prog_gang"), offense-type ("off_mur1", "off_mur2", "off_muratt"), test for adult basic education ("tabe_edu_score"), minimum parole eligibility date (used to generate "years_since_eligible"), and most recent comprehensive risk assessment score ("psych_assess").[103] The first step for this method of extraction was to develop a series of "labeling functions"; a labeling function is a noisy extractor for a task relying on tools such as regular expressions, string searches, or sentiment analysis.[104] Each of our labeling functions includes a prepossessing segmentation function that narrows the text of a long hearing down to one or more smaller chunks of text that are more likely to contain the result. Each labeling function can then either return a result or abstain on the task at the document level.[105]

### 3. *Pre-trained Language Models*

We used pre-trained language models for the following three features: education level ("edu_level"), confirmed job offer ("job_offer"), and year of last disciplinary writeup (used to generate "clean_time"). For these features, the combination of many weakly supervised labeling functions fails to extract the correct value with sufficient accuracy. For such features, we leveraged advances in neural models for information extraction and question

103. Weakly supervised labeling functions are rule-based heuristics (as opposed to probabilistic or stochastic functions) that make a yes/no vote on the label of a data point. Some examples of such functions are keyword search. For example, a function could vote "yes" if the phrase "187 first" appears in the opening statement of the hearing ("187 first" would be of interest because murder is defined in CAL. PENAL CODE § 187 (West 2024)). Multiple such functions are given weights, and then their votes are combined to produce a label for a data point. These labels can then be used for a downstream task called "weak supervision." While supervised learning typically relies on high quality labels that are annotated directly by users, weak supervision relies on lower quality labels, such as the ones generated by the yes/no votes. *See* Alex Ratner, Braden Hancook, Jared Dunnmon, Roger Goldman & Christopher Re, *Snorkel Metal: Weak Supervision for Multi-Task Learning*, *in* DEEM'18: IN'T WORKSHOP ON DATA MGM'T FOR END-TO-END MACH. LEARNING (June 15, 2018).

104. *See* Jenny Hong, Caitlin Voss & Christopher Manning, CHALLENGES FOR INFORMATION EXTRACTION FROM DIALOGUE IN CRIMINAL LAW, *in* PROCEEDINGS OF THE 1ST WORKSHOP ON NLP FOR POSITIVE IMPACT, 73–74 (2021).

105. When combined, multiple labeling functions can comprise a high-quality extractor. We considered several supervised and unsupervised strategies for combining the outputs $\lambda = [\lambda_1, \lambda_2, \ldots]^T$ from the labeling functions into a single label using limited training data. In our exploratory analysis, we found no benefit from using the unsupervised label aggregation models, so we settled on two supervised methods. One used a logistic regression and the other used a constrained least squares model.

answering.[106] We used an approach inspired by the two-step Retriever-Reader approach to open-domain question answering (ODQA).[107]

#### 4. *Reliability of Automated Extraction*

During development, we evaluated the labeling functions on a hold-out development set of documents with manual annotations and data provided by CDCR.[108] We subsequently trained the combination model on a training set. For each task, we computed accuracy statistics on a validation set and chose the more accurate model. All features with an extraction F1 score of below 0.7 were dropped at this stage.

The results of our extraction are summarized in the Appendix. Appendix Table A provides a legend for each feature that met our metrics for reliability and class imbalance. Appendix Table B provides the source and descriptive statistics for each categorical feature. Appendix Table C provides the source and descriptive statistics for each continuous feature.

---

106. *See* Jenny Hong, Derek Chong & Christopher Manning, *Learning from Limited Labels for Long Legal Dialogue*, *presented at* PROCEEDINGS OF THE NATURAL LEGAL LANGUAGE PROCESSING WORKSHOP 2021, 190–204 (2021).

107. *See* Danqi Chen, Adam Fisch, Jason Wetson & Antoine Bordes, *Reading Wikipedia to Answer Open-Domain Questions*, *presented at* PROCEEDINGS OF THE 55TH ANNUAL MEETING OF THE ASSOCIATION FOR COMPUTATIONAL LINGUISTICS, 1870–79 (2017); Rajarshi Das, Shehzaad Dhuliawala, Manzil Zaheer & Andrew McCallum, *Multi-Step Retriever-Reader Interaction for Scalable Open-Domain Question Answering*, *presented at* INTERNATIONAL CONFERENCE ON LEARNING REPRESENTATIONS, 1–13 (2019). Here, our two steps were the Reducer and the Producer. We wrote and trained separate Reducers and Producer for each feature. A Reducer selects relevant segments of text from within a given document. A Producer generates the value of the feature from the reduced text. Using a neural model for the Producer provides many advantages in terms of the complexity of text the model is able to digest. However, most neural models are quite limited in the input length of the text they can handle; many neural models cannot handle more than 500 or 1,000 words at a time, but parole hearings are, on average, 20,000 words. Therefore, we needed a strong Reducer for this context. Using the data produced by the Reducer, we trained a Producer for each feature. We used a pre-trained RoBERTa + BigBird (RoB + BB) base model, which was fine-tuned on various prediction heads. *See* Manzil Zaheer, Guru Guruganesh, Kumar Avinava Dubey, Joshua Ainslie, Chris Alberti, Santiago Ontanon, Philip Pham, Anirudh Ravula, Qifan Wang, Li Yang & Amr Ahmed, *Big Bird: Transformers for Longer Sequences*, *presented at* 34TH CONFERENCE ON NEURAL INFORMATION PROCESSING SYSTEMS, 1–15 (2020). The features "job_offer" and "edu_level" used a sequence classification head. The feature "last_writeup" used a masked language modeling head.

108. To increase the amount of data available for training, validation, and testing beyond the data sources described in the present work, we drew on the annotations produced by prior California parole studies, which were graciously provided by the authors.

## VI. COMPARISON OF MANUAL AND AUTOMATED METHODS

Our overarching research question is methodological: to what extent is machine learning a reliable and useful tool for researching quantitative patterns in parole hearing transcripts? To study this methodological question, we constructed three regression models. Each model was designed to predict whether the Board will grant or deny parole (the outcome variable) based on a large set of factors that have been hypothesized to influence parole decisions (the predictor variables). Each of the three models differ in the source of data for the predictor variables and in the size of the sample.

For the first model, we used data that research assistants manually extracted from transcripts, and data from our public records request. This regression includes thirty-six predictor variables, and the sample size is 688 transcripts. The regression model we built using this manually extracted data reflects the type of quantitative social science research previously done on California parole hearings without the aid of NLP technology.

For the second model, we used predictor variables from tabular data, i.e., data that was provided to us in a structured format and/or that required very little processing. The model includes nine variables provided either directly from CDCR or from the title page or closing page of parole hearing transcripts. In contrast to the manually extracted data, the tabular data was available for the full corpus of hearings (34,993). The regression model we built using tabular data reflects the type of quantitative social science research where there are structured datasets about parole release decisions. The sample size is large, but the number and type of predictor variables is limited.

For the third model, we used predictor variables extracted using the NLP methods discussed in Section V.B. By using NLP, we extracted values for twenty-one predictor variables. This method provides considerably more predictor variables than the tabular data used in the second regression model, and it covers the full corpus of hearings (34,993). There are fewer predictor variables than the manual extraction because we encountered significant challenges in training NLP to extract data. In future work, and with greater investment in the development of NLP technology, we anticipate being able to pull all predictor variables that are in the manual extraction model.

We assessed the effectiveness of the NLP extraction model by comparing the predictive power of the three different regressions. We hypothesized that the NLP model would have greater predictive power than the manual model and the tabular models. Although the NLP model lacks the full set of variables considered in the manual model, it has the advantage of a much larger sample size.

## Table 1: Regressions on the parole outcome.

Adjusted odds ratios (AORs) and Wald $p$-values are reported for all factors in parentheses. Significant values are represented in bold at $p < 0.05$.

| Data Source | (a) Manual | (b) Tabular | (c) NLP |
|---|---|---|---|
| $n$ (Number of Hearings) | 688 | 34,993 | |
| Out-of-Sample AUC | 0.675 | 0.729 | 0.818 |
| **Hearing Actors** | Odds Ratio $e^{\beta}$ ($p$) | | |
| retained_attorney | 1.83 (0.24) | **2.48 (0.00)** | **2.06 (0.00)** |
| commissioner_rate* | **1.14 (0.00)** | **1.34 (0.00)** | **1.39 (0.00)** |
| victim_oppose | **0.13 (0.00)** | - | - |
| victim_present | - | - | **0.43 (0.00)** |
| district_attny_oppose | 0.67 (0.50) | - | - |
| district_attny_present | - | - | **0.68 (0.00)** |
| attorney_opinion | 0.32 (0.08) | - | - |
| **Time & Place** | | | |
| initial_hearing | 0.27 (0.06) | **0.40 (0.00)** | **0.46 (0.00)** |
| years_since_2007 | 1.03 (0.76) | **1.11 (0.00)** | **1.17 (0.00)** |
| years_since_eligible | 1.00 (0.87) | - | 1.00 (0.05) |
| prison_is_level_iv | 1.72 (0.46) | **0.32 (0.00)** | **0.57 (0.00)** |
| **Demographics** | | | |
| ethnicity_black | 1.09 (0.89) | 0.95 (0.17) | 0.95 (0.18) |
| ethnicity_latinx | 1.48 (0.54) | **1.14 (0.00)** | **0.91 (0.02)** |
| ethnicity_other | 0.88 (0.86) | **1.24 (0.00)** | 0.99 (0.88) |
| gender_female | 0.00 (0.99) | **1.23 (0.00)** | **1.28 (0.00)** |
| **Pre-Commitment** | | | |
| justice_involved | 2.27 (0.07) | - | - |
| precommit_sex_abuse | 0.46 (0.24) | - | - |
| precommit_gang | 1.08 (0.90) | - | **1.25 (0.00)** |
| **Commitment Offense** | | | |
| offense_murder_first | 1.46 (0.51) | - | 1.03 (0.45) |
| offense_murder_second | 0.76 (0.62) | - | **1.15 (0.00)** |
| offense_murder_attempt | 0.91 (0.89) | - | **1.12 (0.03)** |
| crime_gang | 1.63 (0.45) | - | - |
| crime_drugs_alcohol | 0.46 (0.18) | - | - |
| claim_innocence | 1.49 (0.54) | - | - |
| **Programs & Rehabilitation** | | | |
| tabe_edu_score | 1.08 (0.75) | - | **1.18 (0.00)** |
| chronos_bucket | **2.16 (0.05)** | - | - |
| programming_gang | - | - | **1.40 (0.00)** |
| programming_all | 1.51 (0.45) | - | - |
| 12steps_program_failed | 0.64 (0.57) | - | - |
| mental_illness | 0.53 (0.12) | - | - |
| mental_treatment | 0.91 (0.88) | - | - |
| **Disciplinary** | | | |
| count_115s | 1.01 (0.66) | - | - |
| clean_time | 1.04 (0.18) | - | **1.02 (0.00)** |
| num_pris_convict_buc | 1.08 (0.94) | - | - |
| **Parole Preparation** | | | |
| psych_assess | **0.45 (0.00)** | - | **0.50 (0.00)** |
| job_offer | 2.21 (0.10) | - | **1.39 (0.00)** |
| **Special Designation** | | | |
| youth_offender | 0.40 (0.17) | - | - |
| elderly_parole | 2.18 (0.25) | - | - |

*For historical commissioner grant rate, AOR is reported in units of a 10% increase in grant rate.

Table 1 presents our three regression models. Regression (a) follows the traditional methodology of analyzing only the set of 688 manually coded transcripts. Regression (b) includes all 34,993 hearings but only analyzes the limited tabular labels provided directly by CDCR as well as features parsed from the hearing title page and concluding page. Finally, regression (c) augments regression (b) with all of the factors extracted using NLP methods.

We calculated the AUC statistic (Area Under the Receiving Operating Curve) under 10-fold cross validation for each model, shown in the second row of Table 1. The manual regression (a) attains an AUC of 0.675. The tabular regression (b) achieves an AUC of 0.729. The NLP regression (c) achieves the highest AUC of the three models: 0.818. When we include the automatically extracted variables, the AUC increases compared to the tabular regression. While its feature set is less comprehensive than that of the manual regression, the massive sample size of the regression with NLP-extracted features enables it to provide the most accurate model for predicting parole decisions.

Having ascertained the reliability of NLP as a tool to extract data and model parole hearings, Parts VII and VIII next investigate the following two sets of hypotheses about California parole hearings:

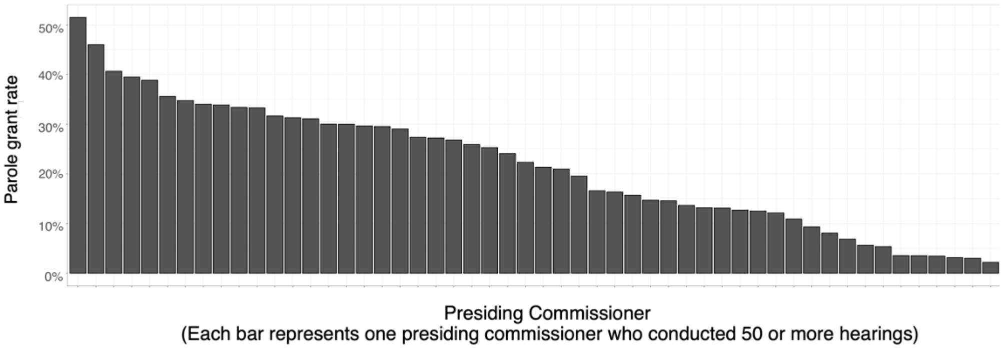
In Part VII, we hypothesized that there would be significant variability in parole release decisions that is attributable to which particular commissioner presided over the hearing.

In Part VIII, we hypothesized that representation by a Board-appointed attorney would be associated with a reduced likelihood of parole. We further hypothesized that this reduction in the likelihood of parole would not be due entirely to selection bias regarding retaining a private attorney. We hypothesized that the reduction in the likelihood of parole would be attributable, at least in part, to differences in what attorneys said at the hearing.

## VII. VARIABILITY BY COMMISSIONER, METHODS AND RESULTS

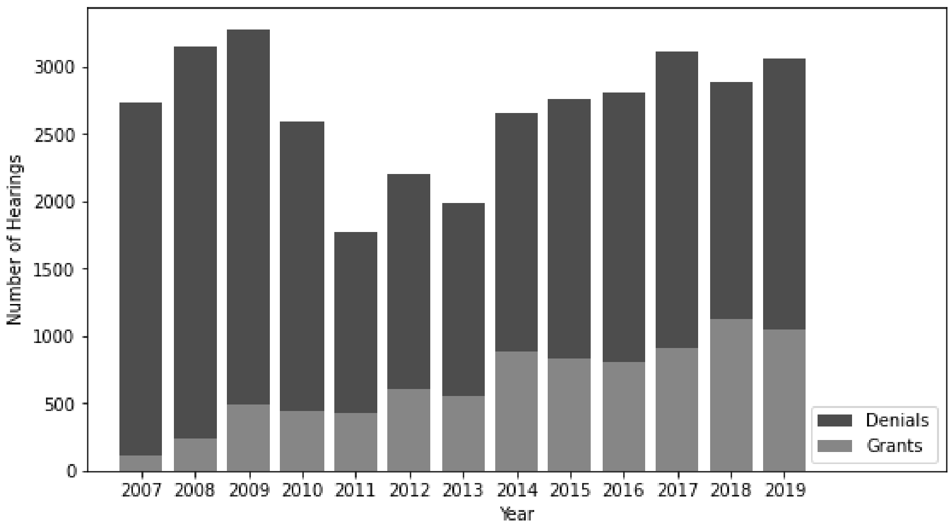
As described in Part V, we extracted the presiding parole commissioner's name, which appears on the title page of each transcript. We then calculated the total number of hearings for each commissioner, and the total number of hearings that resulted in a grant of parole. Figure 1 provides descriptive statistics showing the grant rate for each presiding parole commissioner who conducted fifty or more parole hearings.

**Figure 1: Grant rate for each presiding commissioner**



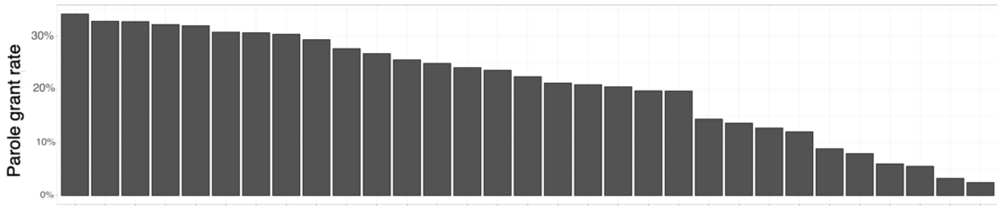

Grant rates by commissioner vary from over 50% to less than 5%, though the cause of this variability is unclear. Commissioners are not randomly assigned across time or across prisons. Figure 2 shows significant variability in parole hearing grant rates by year. Some commissioners served only during the early time period of our study, when grant rates were relatively low. In contrast, others served only during the later time period, when grant rates overall were higher.

**Figure 2: Parole hearing outcomes by year, 2007–2019**




Additionally, some commissioners conduct a higher percentage of their hearings at maximum-security prisons, where grant rates are relatively low. Others conduct the vast majority of their hearings at minimum or medium security prisons, where grant rates are comparatively higher. Figure 3 shows significant variability in grant rates at each prison in which more than fifty parole hearings were held.

**Figure 3: Grant rate at each prison**




Given the variability in grant rates across years and prisons, as seen in Figures 2 and 3, it is reasonable that a given commissioner's average grant rate would implicitly depend on the prisons where they presided and the years in which they served. Consider two hypothetical commissioners who are equal in their decision making. Suppose one commissioner primarily presides over hearings at a medium-security prison, whereas the other primarily presides over hearings at a maximum-security prison. We would expect the latter commissioner to have a lower empirical grant rate, even if there is no underlying difference in their decision-making. Similarly, a commissioner who served only in 2007–2008 would likely have a lower empirical grant rate than a commissioner who served in 2017–2018.

To better understand the independent impact of the commissioner, we used multiple methods: regression analysis and the Monte Carlo randomized inference method.

## A. REGRESSION ANALYSIS

Each regression model in Table 1 includes predictor variables that measure the time period ("year_since_2007"), the prison security level ("prison_is_level_iv"), the variable "commissioner_rate," and several other variables hypothesized to influence parole decisions. "Commissioner_rate" is a continuous variable that measures the grant rate of all prior hearings over which the commissioner presided before the current hearing. Two hearings with the same commissioner can have different values for this variable. "Commissioner_rate" is not collinear with any other variable in the regression.

"Commissioner_rate" is significant in the manual regression, the tabular regression, and the NLP regression. For "commissioner_rate," the adjusted odds ratio (AOR) is reported in units of a 10% increase in grant rate.[109] A commissioner at the 10th percentile of grant rates had, at the time of the hearing, a historical grant rate of 4.76%, while a commissioner at the 90th percentile of grant rates had a rate of 34.37%. Controlling for all other factors, being assigned to a commissioner whose historic grant rate is in the 90th percentile, as opposed to the 10th percentile, is associated with an increase in the odds of parole by 2.7 times in the NLP-based regression and 1.5 times in the manual regression.

As a robustness check, we ran the regression model without "commissioner_rate" and instead included a fixed effect for each individual commissioner. Appendix Table D provides the results; forty of the commissioners are significant in the tabular and NLP regressions at $p<0.05$.

## B. MONTE CARLO RANDOMIZED INFERENCE METHOD

Although commissioners are not assigned to hearings randomly across years and prisons, their assignments within the same prison and within the same year appear to be random.[110] Rather than testing the null hypothesis that all commissioner grant rates are equal across all prisons and years, we tested the following null hypothesis: in a given prison and in a given year, grant rates are independent of the commissioner.

To test commissioner variability, we sampled from this null distribution using a Monte Carlo randomization inference method.[111] Each sample

---

109. An odds ratio is a measure of association between a factor and an outcome (in this case, the grant rate). The odds ratio represents the odds that an outcome will occur given a particular factor, compared to the odds that the outcome will occur in the absence of that factor. An adjusted odds ratio is one that adjusts the odds given additional variables, such as potentially confounding variables. Logistic regression is one such method for adjusting odds ratios. *See, e.g.*, Magdalena Szumilas, *Explaining Odds Ratios*, 19 J. CAN. ACAD. CHILD ADOLESCENT PSYCHIATRY 227 (2010).

110. *See* Laqueur & Venancio, *supra* note 9, at 215 (conducting a randomization and finding that, within a given prison and a given year, decision-makers are randomly assigned).

111. *See* Abrams et al., *supra* note 23, at 359–60 (using Monte Carlo method). In order to test statistical hypotheses, we must determine whether a single numerical measurement (in this case, a commissioner's grant rate) comes from the null distribution or not (the case where all commissioners have the same underlying grant rate given the same case). The null distribution is not a statistical distribution with a well-known form like the Normal distribution, the Binomial distribution, etc. Rather, the main way that we can measure our particular null distribution is through Monte Carlo sampling. In other words, we sample a large number of points from the distribution until we can form its approximate shape. Randomization inference is a type of statistical inference that holds a certain property as an invariant, and randomizes all other properties. In our scenario, we assumed that the year and prison where a hearing is held is fixed, but that there is some possibility of a hearing being assigned to one

represents a new assignment of commissioners to hearings, that is, a new list of 34,993 commissioner-hearing pairs. We calculated each commissioner's grant rate for each assignment sample, excluding those who have presided over fewer than fifty hearings during their tenure. We considered 10,000 assignments sampled under the null distribution for each of the fifty-two commissioners and the resulting grant rates under this null distribution.

For each commissioner, we computed a 0.05%–99.95% interval of the commissioner's grant rate under the null distribution. This means that any individual commissioner's grant rate that falls outside the range does so with a probability of 0.001.

Using a Bonferroni correction for the fifty-two rates under consideration, we rejected the null hypothesis across commissioners at a familywise error rate of $\alpha = 0.05$.[112] Fourteen of the fifty-two commissioners fall outside their Bonferroni-corrected intervals at this $\alpha = 0.05$ level, (i.e., they have higher or lower grant rates than expected, assuming that prison and year are the sources of variance in a commissioner's grant rate).

## VIII. VARIABILITY BY ATTORNEY TYPE, METHODS, AND RESULTS

The variable "retained_attorney" in Table 1 indicates whether a person privately retained an attorney. Retained attorneys include attorneys who were hired and paid for by parole candidates (and/or their families and friends), as well as attorneys who provided pro bono representation. Nearly all parole candidates who did not retain a private attorney were represented by a Board-appointed attorney, who was appointed and paid for by the Board.[113] Figure 4 provides descriptive statistics about attorney-type.

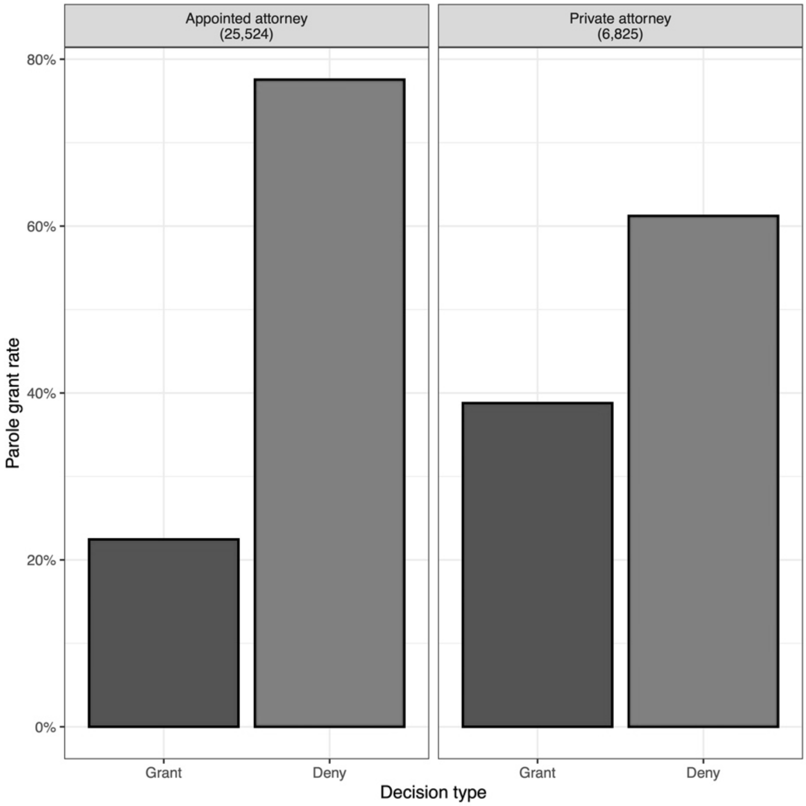
---

commissioner (who also presided over hearings at that prison in that year) compared to another. *See* David M. Ritzwoller, Joseph P. Romano & Azeem M. Shaikh, *Randomization Inference: Theory and Applications*, ARXIV (June 2024), https://arxiv.org/abs/2406.09521.

112. The Bonferroni correction is an adjustment to the p-value of a statistical test to correct for the error rate across multiple tests. To understand the intuition behind the Bonferroni correction, consider a situation where the same statistical test is run twenty times. By sheer chance alone, we would expect the p-value to fall below 0.05 in one of these experiments. So we cannot simply use the test for whether a *single* p-value falls below 0.05, we want to determine how likely it is that *all twenty tests together* show a result outside of the null distribution. The Bonferroni correction is a correction that divides the individual p-values by the number of experiments. *See* Gareth James, Daniela Witten, Trevor Hastie, Robert Tibshirani & Jonathan Taylor, AN INTRODUCTION TO STATISTICAL LEARNING WITH APPLICATIONS IN PYTHON 567–68 (2023).

113. A small number of individuals declined to be represented by an attorney and chose instead to represent themselves. Because these individuals did not retain private attorneys, we categorized these individuals as not having retained private attorneys.

**Figure 4: Grants and denials by attorney type**




The regression analysis in Table 1 shows that representation by a retained attorney is associated with a significant increase in the likelihood of being granted parole. In the NLP regression, the odds of being granted parole for an average candidate are 2.1 times higher for those with retained attorneys. As discussed above, another study of California parole hearings made a similar finding, but it was limited in its ability to draw a causal connection between retaining a private attorney and increased odds of release.[114] The study noted that retaining a private attorney may act as a proxy for a higher socio-economic class.[115] Except for those who receive pro bono private counsel, only those candidates who have sufficient economic means (or whose families have sufficient economic means) can afford to retain a private attorney.[116] Given research that shows lower socio-economic class is associated with increased

114. *See* Bell, *supra* note 4, at 488.

115. *See id.* at 488 n.145.

116. *See* CAL. CODE REGS. tit. 15, § 2256 (2023) (prisoner presumed unable to afford attorney if they have less than $1,500 or they show that they are unable to retain an attorney for that amount).

incarceration,[117] it is plausible to hypothesize that a candidate's lower socio-economic class could reduce the likelihood of parole. If that were the case, then one would expect the variable "retained_attorney" to increase the odds of parole in Table 1, even if the quality of retained attorneys were equal to that of Board-appointed attorneys. We have no measure for socio-economic class in the dataset, so it is challenging to isolate the relative quality of representation provided by retained versus appointed attorneys and discern whether there is a difference in quality that has any causal impact on parole outcomes.

To approach this challenge, we first conducted a regression analysis to predict whether a candidate will retain private legal representation. Appendix Table E provides the results of that analysis. The outcome variable is whether a candidate retained a private attorney for a hearing. Predictor variables include a subset of variables that are hypothesized to influence parole decisions. The subset consists of all variables from Table 1 aside from those that a candidate may not know about before they must decide whether to retain a private attorney. For example, variables like the rate at which the commissioner has granted parole, the risk assessment score, and whether the district attorney attends the hearing are excluded because they may be unknown when a candidate decides to retain an attorney.[118] We restricted the analysis to the set of transcripts where the attorney type is known (n = 32,349).

Several variables have a significant effect on both the decision to retain a private attorney, as seen in Appendix Table E, and on the likelihood of parole as seen in Table 1. These include but are not limited to: those with a confirmed job offer (have higher odds of retaining an attorney and being granted parole); gender (female candidates have higher odds of retaining an attorney and being granted parole); and ethnicity (Latinx candidates have lower odds of retaining an attorney and being granted parole). This overlap of significant variables in Table 1 and Appendix Table E underscores the difficulty in investigating why retaining a private attorney is associated with an increased likelihood of parole. The increase may be explained by the fact that factors which make a person more likely to be granted parole are also factors which make a person more likely to retain an attorney. How can we ascertain whether the attorney's time or skill in preparing the candidate for the hearing or what the attorney is

117. *See, e.g.*, Bruce Western & Becky Pettit, *Incarceration and Social Inequality*, 139 DAEDALUS 8 (2010).

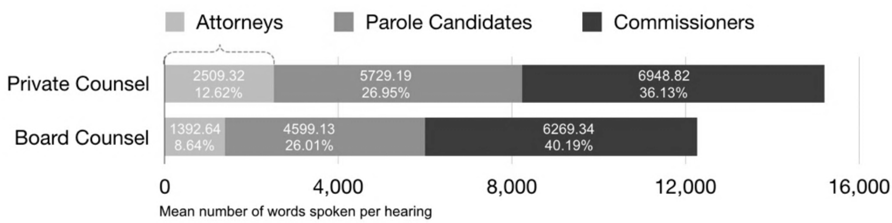
118. The Board appoints attorneys 120 days prior to the hearing. *See* MacKay, *supra* note 84, at 288. Candidates who retain attorneys generally do so many months, or years, before their hearing in order to have adequate time to prepare for the hearing. The district attorney from the county of conviction receives information about the candidate approximately sixty days before the hearing. *See id.* at 285. Candidates receive a copy of the comprehensive risk assessment sixty days before the hearing. *See id.* at 284.

actually saying (or not saying) in the hearing makes any causal difference in the outcome? We cannot fully answer this question because, in addition to the limits of observational data, we have no measure of how much time attorneys spend preparing their clients. Leveraging NLP does, however, allow us to investigate the relationship between the likelihood of parole and what an attorney said (or did not say) in the hearing itself.

We conducted multiple linguistic analyses that provide different insights into how privately retained and Board-appointed attorneys speak at hearings. Taken together, these analyses support the hypothesis that the reduction in the likelihood of parole that is associated with a Board-appointed attorney is attributable, at least in part, to differences in what the attorneys said at the hearing.

Figure 5 shows that speaking times for the parole candidate, attorney, and commissioner differ significantly between hearings in which representation is provided by a privately retained attorney compared to a Board-appointed attorney. We measured an individual's speaking time by counting the number of words that are attributed to them in the hearing transcript. The mean speaking time for privately retained attorneys is 80% longer than the mean speaking time for Board-appointed attorneys. Parole candidates speak 25% more when represented by a privately retained attorney. These differences in speaking times are statistically significant ($p < 0.05$) in a Chi-squared test. The next two methods investigate the particular words that retained attorneys tend to use in comparison to Board-appointed attorneys, and whether those particular words are associated with increased likelihood of parole.

**Figure 5: Speaking time breakdown by attorney status** (all differences statistically significant ($p < 0.05$) in a Chi-squared test)

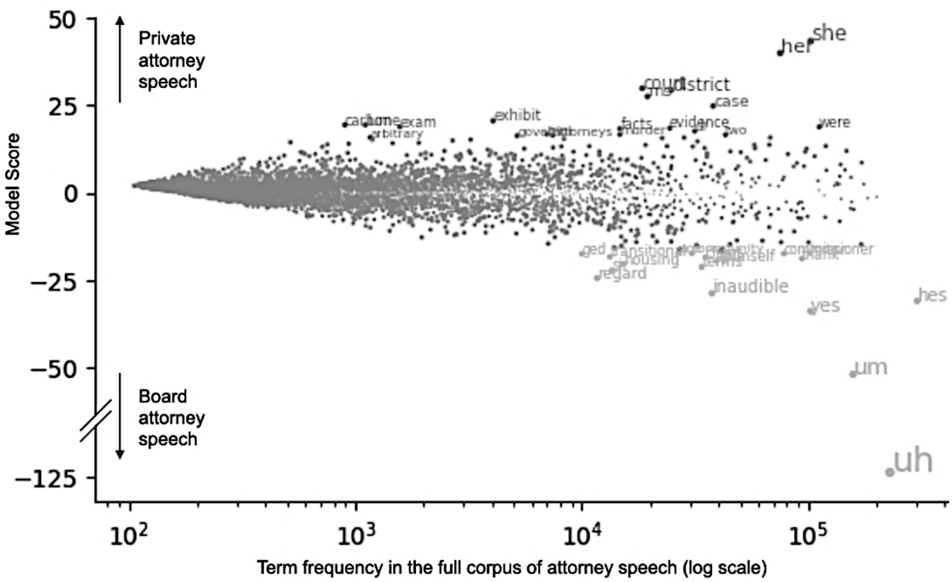


### A. WORD POLARITY SCORE

Using the model-based word score method of Monroe, Colaresi, and Quinn,[119] derived from normalized log odd ratios,[120] we identified the most polar words that explain the difference between privately retained and Board-appointed attorney speech. Appendix F describes the method used to calculate a polarity score for each word used in the corpus. A word's polarity score measures how effectively that word distinguishes between privately retained attorneys and Board-appointed attorneys.

Figure 6 shows word polarity scores plotted against occurrence frequency. The top ten words most indicative of privately retained attorney speech largely include legal terms such as "court," "case," and "evidence" (as well as female pronouns[121]). "Uh," "um," "yes," and "sir" are among the top ten words indicative of Board-appointed attorney speech.

**Figure 6: Word polarity scores in attorney speech**




We tested whether these words are individually associated with a higher or lower likelihood of a decision to grant parole. We conducted a regression

119. *See* Burt L. Monroe, Michael P. Colaresi & Kevin M. Quinn, *Fightin' Words: Lexical Feature Selection and Evaluation for Identifying the Content of Political Conflict*, 16 POL. ANALYSIS 372 (2008).

120. Log odd ratios are the natural logarithm of the odds ratios. *See supra* Section V.B.

121. The fact that female pronouns are more common in the language of private attorneys is consistent with the fact that female candidates are significantly more likely to retain private counsel. *See* Appendix Table E.

analysis where the outcome variable is the parole-decision and the predictor variables include the frequency count of the top ten words most indicative of privately retained attorney and Board-appointed attorney speech, attorney speaking time, and all features from the NLP regression reported in Table 1. Several of the top ten privately retained attorney words are significant ($p<0.01$) and are associated with increased odds of parole; these words include "district," "court," and "exhibit." The following top ten Board-appointed attorney words are significant ($p<0.01$) and are associated with decreased odds of parole: "uh," "um," "yes," "sir," and "inaudible."

### B. LEGAL LEXICON

The word polarity analysis is unguided in the sense that it does not include direction about which words to find. We supplemented this unguided approach with analysis that is informed by specific background knowledge about California parole hearings. Following the approach of others who do computational analysis of text, we curated a lexicon to investigate the attorney's usage of specific legal language.[122] We consulted a legal expert to generate a list of terms and phrases that they hypothesized a skilled attorney might be likely to use in representing clients at a parole hearing. The expert is a member of our research team who has experience representing a parole candidate at a California parole hearing, appealing parole release decisions, and doing prior legal research on California parole hearings. For each term and phrase, we computed the percentage of hearings in which the term or phrase is used by the attorney at least once. Several terms and phrases were dropped because they appeared in five percent or less of transcripts.[123] The following eleven terms and phrases remained: references to case names regarding parole ("Lawrence,"[124] "Shaputis,"[125] "v," and "in re"), key phrases drawn from statutes and caselaw on parole ("some evidence,"[126] "unreasonable risk to

122. *See, e.g.*, Dallas Card, Serina Chang, Chris Becker, Julia Mendelsohn, Rob Voigt, Leah Boustan, Ran Abramitzky & Dan Jurafsky, *Computational Analysis of 140 Years of US Political Speeches Reveals More Positive but Increasingly Polarized Framing of Immigration*, 119 PNAS 7 (2022) (using curated lexicon of immigration terms).

123. Terms that appeared in less than five percent of transcripts include several cases ("Rosencrantz," "In re Butler," and "Miller v. Alabama"), phrases from statutes and caselaw ("current danger," "diminished culpability," "transient"), and "reasonable doubt."

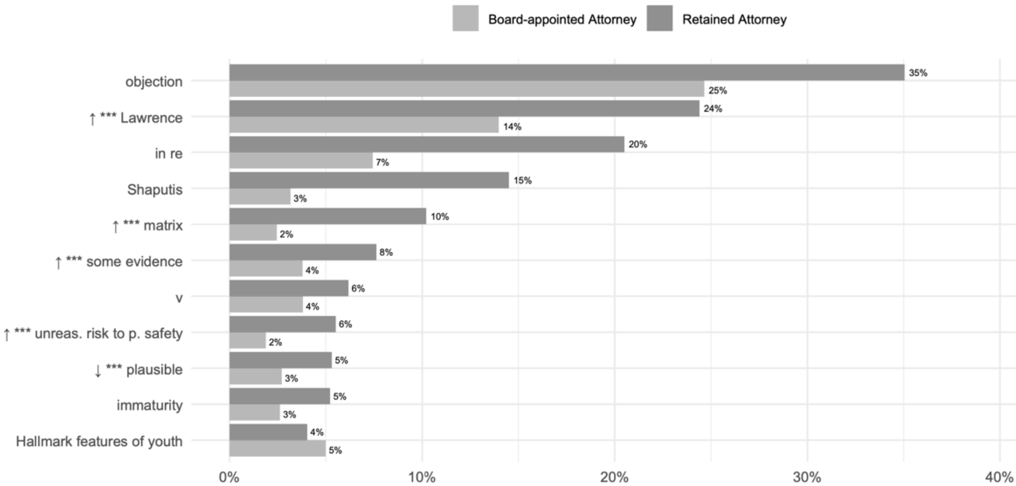
124. *See In re* Lawrence, 44 Cal. 4th 1181 (2008).

125. *See In re* Shaputis, 44 Cal. 4th 1241 (2008).

126. *See Lawrence*, 44 Cal. 4th at 1181 ("some evidence" used throughout the opinion).

public safety,"[127] "matrix,"[128] "plausible,"[129] "hallmark features of youth,"[130] "immaturity"[131]), and "objection." For these terms and phrases, we ran t-tests to determine whether there are statistically significant differences in the respective rates at which Board-appointed attorneys and privately retained attorneys use these words and phrases. As shown in Figure 7, Board-appointed attorneys use ten of the eleven terms and phrases significantly less at a $p < 0.05$ threshold.

**Figure 7: Term frequency in attorney speech**




Legal term usage by retained vs. appointed attorneys. Terms that are significant in a regression analysis that includes all variables in Table 1 are marked with *** for p<0.001, ** for p<0.01, * for p<0.05. For each term that is marked as significant, an arrow indicates whether mention of that term increases (↑) or decreases (↓) the odds of a parole grant.

Figure 7 also provides results from a regression analysis that includes these terms and phrases as well as all other predictor variables considered in the NLP regression in Table 1. The following five of the eleven terms and phrases significantly predict the parole outcome at a $p < 0.001$ level: "Lawrence," "matrix," "some evidence," "unreasonable risk to public safety," and "plausible." Each of these terms and phrases increases the probability of a

127. *See id.* at 1221 (citing 15 Cal. Code. Regs. Section 2281(a)).

128. *See In re* Dannenberg, 34 Cal. 4th 1061, 1078–79 (2005) (citing 15 Cal Code Regs. § 2403(a) and explaining the role matrix played in parole hearings as of 2005); *In re* Butler, 4 Cal. 5th 728, 734 (2018) (explaining that, as of 2016, matrix no longer plays defined role in parole hearings due to changes in law).

129. *See In re* Shaputis II, 53 Cal. 4th 192, 216 (2011).

130. *See* CAL. PENAL CODE § 3051(f) (West 2024).

131. *See id.* (referencing "maturity").

parole grant, except for "plausible" which decreases the probability of a parole grant.[132]

## IX. DISCUSSION, LIMITATIONS, AND FUTURE RESEARCH

Parole is the heavy gate at the end of the criminal legal system's long corridor. Its keepers can tip a sentence from fifteen years to fifty. Many of the mechanisms underlying parole have remained opaque not only to the public, but also to governmental oversight bodies. Leveraging the unstructured data recorded in parole hearing transcripts, we employed machine learning tools to extract and analyze parole case factors and shine a light onto this system.

We found that automated extraction provides a reliable method for extracting data from text. Automated extraction allows for a much larger sample size than prior studies of parole hearings. Moreover, we showed in Table 1 that a regression model that includes automatically extracted data over this large sample size has a better predictive fit than regression models using only manually extracted data or tabular data.

While our findings indicate that NLP is an effective tool for studying parole hearings, recent advances could unlock even more effective analyses. We were limited by the sophistication of NLP tools available to us at the time when we were conducting our research. Given those limitations, we were not able to automatically extract all features of interest with sufficient reliability. Each feature presented unique challenges. Some features were more difficult because they were relatively nuanced and required some judgment. For example, similar to prior studies, we manually annotated whether a parole candidate was asked questions about the Twelve Steps of Alcoholics/ Narcotics Anonymous and whether the commissioner took the response to be adequate or a failure.[133] The description of this variable in the coding manual is considerably detailed, including five examples of what constitutes an adequate response and several clarifications to questions raised by research assistants who did annotations. The training for research assistants focused more on this type of variable as compared to straightforward variables such as the commitment offense and whether it was the candidate's initial hearing

132. An initial hypothesis for why "plausible" decreases the likelihood of parole is that this word is used in a case that explains the rights of candidates who maintain their innocence. *See Shaputis II*, 53 Cal. 4th at 216. Although we did not find that claiming innocence reduces the likelihood of parole, others have suggested that claiming innocence does make it more difficult to be granted parole. *See, e.g.*, Daniel S. Medwed, *The Innocent Prisoner's Dilemma: Consequences of Failing to Admit Guilt at Parole Hearings*, 93 IOWA L. REV. 491, 529 (2008). More research is needed in this area.

133. *See* Young et al., *supra* note 31; Bell, *supra* note 4.

before the Board. A feature that requires more nuance for humans to reliably annotate is likewise difficult for NLP to reliably extract.

Other features were especially difficult to extract because they appeared in different parts of the transcripts. For example, discussion about whether a candidate had been treated for mental illness sometimes appeared during the initial part of the transcript, sometimes during discussion of social history and the crime, and sometimes during discussion of a person's time in prison. The fact that this type of feature is not generally confined to a specific part of the transcript makes it challenging to identify the relevant part of the document in which to look for it. Our NLP models were relatively limited in the amount of text that they could ingest at a given time. The average transcript was 18,499 words and, at the time we did the analysis, NLP tools were not able to take in that amount of text in one pass. We used various document segmentation methods such as topic modeling via latent Dirichlet allocation, but those ultimately failed to produce any semblance of meaningful segments. We are optimistic, however, because retrieval-augmented language models are an area of active research and these could help solve this type of challenge. State-of-the-art large language models (LLMs) are constantly pushing the boundaries of long text.

Another challenge we found was that our pretrained language model could perform our extraction tasks only via supervised learning, meaning that we needed to give it many training labels. This approach required a large amount of quality training data, which is labor-intensive to obtain. In contrast, the general-purpose LLMs of 2025 can perform some tasks with no additional task-specific training data, a technique that is referred to as zero-shot learning. Challenges would likely persist with especially nuanced variables, but specific training data could be incorporated to do few-shot learning in regard to these variables.

An important aspect of LLMs to consider before using them, however, is control over sensitive data. We did not use LLMs in this work because at the time we did our analysis, the most powerful LLMs were available to us only through API access, and data could be shared with the model providers.[134] This kind of sharing would have raised ethical concerns about data privacy. In the near future, however, it will likely be possible to run a sufficiently powerful LLM in a way that does not share data.[135] For these reasons, we are optimistic

134. *See* Amy Winograd, *Loose-Lipped Large Language Models Spill Your Secrets: The Privacy Implications of Large Language Models*, 36 HARV. J. L. & TECH. 615, 623 (2023).

135. *See, e.g.*, S. Liu et al., *RTLCoder: Fully Open-Source and Efficient LLM-Assisted RTL Code Generation Technique*, IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems (2024).

that recent rapid advances in NLP technology will allow future studies to extract all additional features of interest. Thus, the power of NLP parole hearing models will only increase with time and the development of NLP technology.

In addition to our general conclusion that NLP offers a promising avenue for research of legal decision-making, we also made specific findings about parole hearings in California. The NLP regression in Table 1 shows that a presiding commissioner whose historic grant rate is in the 90th percentile as opposed to the 10th percentile is associated with an increase in the odds of parole by 2.7 times. Further, privately retaining an attorney is associated with an increase in the odds of parole by 2.1 times. These findings are noteworthy when compared to case factors that many argue should make a difference to parole outcomes as a normative matter. For example, the regression analysis in Table 1 shows that completing four or more types of rehabilitation programs is not significant in predicting the likelihood of parole.[136] The number of years since the last write-up for prison misconduct is significant, but each year without a write-up is associated with increasing the odds of parole by 1.02 times. The effect size of this variable is smaller than either commissioner grant rate or retained attorney.

Notably, several factors that are within the discretion of other actors are significant in the regression analysis. If a victim representative or the district attorney appears at the hearing (in almost all cases to make a statement opposing parole), each of them is associated with the odds of a parole grant dropping by 2.3 times and 1.5 times respectively.[137] This finding deserves a closer look in future research. Further, the odds of parole are 4.0 times lower for a candidate who receives a "high risk" score as compared to a "low risk" score on a risk assessment conducted by one of the Board's forensic psychologists. Receiving a "moderate risk" score as compared to a "low risk" score is associated with a reduction in the odds of parole by 2.0 times.

136. The variable "programming_all," which is included in the manual regression and is not statistically significant, measures whether a candidate participated in four or more types of rehabilitation programs. *See* Appendix A. The NLP regression does not include "programming_all" because our NLP methods were not successful in reliably extracting all types of rehabilitation programs. Our NLP methods did reliably extract one type of rehabilitation program, namely those focused on gang membership, and the NLP regression therefore includes the variable "programming_gang." Participation in a gang-focused program is associated with increasing the odds of parole by 1.4 times.

137. The NLP-extracted regression uses victim and district attorney presence, respectively, because the machine learning classifier for opposition was unable to outperform a baseline for presence. *See supra* note 102. In the manual regression, we use the more granular variable.

We were careful not to draw causal inferences from the results of our regression analysis in Table 1 and we used additional methods to isolate variability introduced by commissioners and attorneys. We measured commissioner variability in the presence of non-random assignment to hearings and found significant excess variability in grant rates beyond what should be expected. This finding aligns with prior research on California parole hearings, which also found significant variability among commissioners.[138] We also conducted linguistic analysis on the words used by Board-appointed attorneys as compared to retained attorneys. We found that retained attorneys use significantly more sophisticated legal language, and that this increased sophistication in language is associated with a higher likelihood of parole.

## X. CONCLUSION

This Article has presented our initial reconnaissance findings about parole release decisions with the aid of NLP. In future work, we aim to develop a reconsideration tool to identify anomalous parole denials for second-look review by an oversight body or appellate attorneys. The purpose of creating a list of denials for second-look review is to provide opportunities to improve fairness in discretion, so it is critical that the list itself is generated equitably.

In our earlier work, we explained our concern about constructing a reconsideration tool without first doing reconnaissance.[139] We considered the possibility of a "reconsideration-only" tool that does not describe patterns in the existing set of decisions, but simply identifies outliers. Outliers could be given to a second-look body for possible reconsideration, then data about whether the body ultimately changed the decisions could be inputted as feedback. The tool could be programmed to improve on the task of finding decisions that are likely to be altered upon reconsideration. Our findings in this Article show the danger of building such a tool without first adequately understanding existing patterns in decisions. We now know that hearings with Board-appointed attorneys and presiding commissioners with historically lower grant rates are more likely to be associated with the denial of parole. A denial at such a hearing would, all other things equal, be less likely to be flagged as anomalous than a denial at a hearing with a retained attorney and a commissioner who has a historically higher grant rate. A parole candidate would be less likely to be put on the list for second-look review in virtue of the very features that suggest potential inequity in the exercise of discretion in their case. Such a tool would reify rather than rectify potential inequity.

138. *See* Laqueur & Venancio, *supra* note 9, at 216–17.

139. *See* Bell, *supra* note 2, at 833–34.

What would it look like to develop a reconsideration tool that is informed by reconnaissance findings? A full answer in the context of California parole must be informed by more thorough reconnaissance findings, especially findings about race and ethnicity. Other research has found racial disparity in parole release decisions, and we plan to investigate the issue in future work.[140] While a full answer is not yet available, we can describe what it would look like to build a reconsideration tool that is informed by the initial reconnaissance findings here. Such a tool could flag denials as anomalous in cases where a regression model predicts that a candidate would have been granted parole if they had been judged by a different commissioner and/or retained a private attorney. If changing one or both of these variables would change the predicted outcome, the decision could be categorized as "high risk" for depending on features that should make no normative difference to the outcome. These decisions would form the start of a list for second-look review, and more decisions would be added based on additional reconnaissance findings.

There is no guarantee that a "high risk" decision would change on a second-look review. Just as many people who are "high risk" for recidivism on a predictive algorithm do not re-offend,[141] many decisions categorized as "high risk" for inequity on a reconsideration tool may be fully justified. Either way, we believe it is myopic to focus technological development on predicting the behavior of parole candidates and others who are processed through the criminal legal system. It is high time to focus research and attention on technological tools designed to scrutinize the exercise of discretion in the criminal legal system.

---

140. *See* Bell, *supra* note 4, at 460.

141. *See, e.g.*, Marcus T. Boccaccini, Daniel C. Murrie, Jennifer D. Caperton & Samuel W. Hawes, Field Validity of the STATIC-99 and MnSOST-R Among Sex Offenders Evaluated for Civil Commitment as Sexually Violent Predators, 15 PSYCH. PUB. POL'Y & L. 278, 302 (2009) (finding 6.3% recidivism rate among individuals categorized as high-risk on the STATIC-99 risk assessment tool).

# XI. APPENDIX

## A. LEGEND OF VARIABLES

| Variable | Description |
|---|---|
| **Hearing Actors** | |
| retained_attorney | Whether the candidate privately engaged an attorney |
| commissioner_rate | Historical grant rate of the presiding commissioner at the time of the hearing |
| victim_oppose | Does the victim make a statement opposing parole? (Used in the manual regression but not the NLP regression.) |
| victim_present | Victim present at hearing? (Used in the NLP regression in place of victim_oppose.) |
| district_attny_oppose | Does the DA make a statement opposing parole? (Used in the manual regression but not the NLP regression.) |
| district_attny_present | Is the DA present at the hearing? (Used in the NLP regression in place of district_attny_oppose.) |
| attorney_opinion | In the closing statement did the candidate's attorney argue for release? |
| **Time & Place** | |
| initial_hearing | Is this the candidate's first hearing? |
| years_since_2007 | Year of the hearing (since the first year of the dataset) |
| years_since_eligible | Number of years candidate has served over their lowest applicable parole eligibility date |
| **Demographics** | |
| ethnicity_black | CDCR-recorded ethnicity = "Black" |
| ethnicity_latinx | CDCR-recorded ethnicity = "Hispanic/Latino" |
| ethnicity_other | CDCR-recorded ethnicity = "Other" |
| ethnicity_white | CDCR-recorded ethnicity = "White" |
| gender_female | CDCR-recorded gender = female (indicated by CDCR number beginning with the letter W) |
| **Pre-Commitment** | |

| | |
|---|---|
| justice_involved | Did candidate have prior convictions, prior parole or probation, or prior incarceration? |
| precommit_sex_abuse | Victim of sexual abuse prior to commitment offense? |
| precommit_gang | Whether person was involved in gang activity prior to commitment offense |
| **Commitment Offense** | |
| offense_murder_first | At least one count of murder in the first-degree |
| offense_murder_second | At least one count of murder in the second-degree |
| offense_murder_attempt | At least one count of of attempted murder |
| crime_gang | Was crime rooted in gang activity? |
| crime_drugs_alcohol | Whether person was intoxicated at time of crime or heavily using alcohol/drugs around the time |
| claim_innocence | Does Candidate claim innocence in the commitment offense? |
| **Programs & Rehabilitation** | |
| tabe_edu_score | Most recent score on TABE (test for adult basic education), histogram-bucketed into these categories 0 (score of 0–8.9); 1 (score of 9–11.9); or 2 (score of 12+) |
| chronos_bucket | Number of positive write-ups from prison staff ("laudatory chronos"), histogram-bucketed into these categories 0 (none); 1 (1–9 laudatory chronos); 2 (10+ laudatory chronos) |
| programming_all | Participated in programs in at least 4 of the following categories: anger management, art or fitness, education, gang, parenting, philanthropic, religious, substance abuse, therapy, victim, vocaction, other |
| programming_gang | Participated in gang programming (Used in the NLP regression, since it was the only programming variable that we reliably extracted; in the manual regression, this is included as part of programming_all.) |

| | |
|---|---|
| 12steps_program_failed | Whether candidate was asked about the Twelve Steps and did not give an adequate response |
| mental_illness | History of diagnosed mental illness? |
| mental_treatment | Currently receiving mental health treatment (medication or counseling)? |
| **Disciplinary** | |
| count_115s | Total count of 115s (disciplinary writeup forms) |
| clean_time | Years since last disciplinary infraction (infraction defined as 115 in the NLP regression; infraction defined as 115 or prison conviction in the manual regression) |
| num_pris_convict_buc | Number of convictions while in prison, histogram-bucketed into categories 0 (no prison convictions); 1 (one or more prison convictions) |
| prison_is_level_iv | Whether hearing took place at a prison where more than half of the population is level IV |
| **Parole Preparation** | |
| psych_assess | Score at most recent comprehensive risk assessment, bucketed into categories 0 (low or low/moderate); 1 (moderate); 2 (moderate/high or high) |
| job_offer | Confirmed job offer? |
| **Special Designation** | |
| youth_offender | Youth Offender Parole Hearing |
| elderly_parole | Elderly Parole Hearing |

### B. Source and Descriptive Statistics for Categorical Variables

| Variable name | Source | Total n | Decision grant n | Decision grant percentage |
|---|---|---|---|---|
| **Hearing Actors** | | | | |
| retained_attorney (0) | NLP/ tabular | 25524 | 5730 | 22.4 |

| | | | | |
|---|---|---|---|---|
| retained_attorney (1) | | 6825 | 2647 | 38.8 |
| victim_oppose (0) | Manual | 573 | 151 | 26.4 |
| victim_oppose (1) | | 110 | 17 | 15.5 |
| victim_present (0) | NLP/ tabular | 30102 | 7504 | 24.9 |
| victim_present (1) | | 4891 | 986 | 20.2 |
| district_attny_oppose (0) | Manual | 72 | 37 | 51.4 |
| district_attny_oppose (1) | | 610 | 130 | 21.3 |
| district_attny_present (0) | NLP/ tabluar | 3282 | 781 | 23.8 |
| district_attny_present (1) | | 31711 | 7709 | 24.3 |
| attorney_opinion (0) | Manual | 521 | 152 | 29.2 |
| attorney_opinion (1) | | 167 | 17 | 10.2 |
| **Time & Place** | | | | |
| initial_hearing (0) | NLP/ tabular | 27945 | 7397 | 25.5 |
| initial_hearing (1) | | 7048 | 1093 | 15.5 |
| **Demographics** | | | | |

| | | | | |
|---|---|---|---|---|
| ethnicity_black (0) | NLP/ tabular | 23345 | 5875 | 25.2 |
| ethnicity_black (1) | | 11643 | 2614 | 22.5 |
| ethnicity_latinx (0) | NLP/ tabluar | 24779 | 5977 | 24.1 |
| ethnicity_latinx (1) | | 10209 | 2512 | 24.6 |
| ethnicity_other (0) | NLP/ tabular | 31082 | 7397 | 23.8 |
| ethnicity_other (1) | | 3906 | 1092 | 28 |
| ethnicity_white (0) | NLP/ tabular | 25758 | 6218 | 24.1 |
| ethnicity_white (1) | | 9230 | 2271 | 24.6 |
| gender_female (0) | NLP/ tabular | 33262 | 7940 | 23.9 |
| gender_female (1) | | 1731 | 550 | 31.8 |
| **Pre-Commitment** | | | | |
| justice_involved (0) | Manual | 236 | 51 | 21.6 |
| justice_involved (1) | | 452 | 118 | 26.1 |
| precommit_sex_abuse (0) | Manual | 613 | 151 | 24.6 |
| precommit_sex_abuse (1) | | 75 | 18 | 24 |
| precommit_gang (0) | NLP/ tabular | 19556 | 4255 | 21.8 |

| | | | | |
|---|---|---|---|---|
| precommit_gang (1) | | 15437 | 4235 | 27.4 |
| **Commitment Offense** | | | | |
| offense_murder_first (0) | NLP/ tabular | 8718 | 1963 | 22.5 |
| offense_murder_first (1) | | 26275 | 6527 | 24.8 |
| offense_murder_second (0) | NLP/ tabular | 21839 | 5220 | 23.9 |
| offense_murder_second (1) | | 13154 | 3270 | 24.9 |
| offense_murder_attempt (0) | NLP/ tabular | 31458 | 7668 | 24.4 |
| offense_murder_attempt (1) | | 3535 | 822 | 23.3 |
| offense_gang (0) | Manual | 554 | 132 | 23.8 |
| offense_gang (1) | | 114 | 114 | 28.9 |
| crime_drugs_alcohol (0) | Manual | 113 | 25 | 22.1 |
| crime_drugs_alcohol (1) | | 575 | 144 | 25 |
| claim_innocence (0) | Manual | 592 | 153 | 25.8 |
| claim_innocence (1) | | 96 | 16 | 16.7 |
| **Programs & Rehabilitation** | | | | |

| | | | | |
|---|---|---|---|---|
| tabe_edu_score (0) | NLP/ tabular | 8747 | 1798 | 20.6 |
| tabe_edu_score (1) | | 7161 | 1829 | 25.5 |
| tabe_edu_score (2) | | 10735 | 3138 | 29.2 |
| chronos_bucket (0) | Manual | 319 | 55 | 17.3 |
| chronos_bucket (1) | | 337 | 102 | 30.3 |
| chronos_bucket (2) | | 31 | 12 | 38.7 |
| programming_all (0) | Manual | 159 | 22 | 13.8 |
| programming_all (1) | | 529 | 147 | 27.8 |
| programming_gang (0) | NLP/ tabular | 26685 | 5477 | 20.5 |
| programming_gang (1) | | 8308 | 3013 | 36.3 |
| 12steps_program_failed (0) | Manual | 624 | 163 | 26.1 |
| 12steps_program_failed (1) | | 64 | 6 | 9.38 |
| mental_illness (0) | Manual | 314 | 89 | 28.3 |
| mental_illness (1) | | 371 | 78 | 21 |
| mental_treatment (0) | Manual | 570 | 145 | 25.4 |
| mental_treatment (1) | | 118 | 24 | 20.3 |

| | | | | |
|---|---|---|---|---|
| **Disciplinary** | | | | |
| num_pris_convict_buc (0) | Manual | 637 | 161 | 25.3 |
| num_pris_convict_buc (1) | | 51 | 8 | 15.7 |
| prison_is_level_iv (0) | NLP/ Tabular | 30592 | 8079 | 26.4 |
| prison_is_level_iv (1) | | 4401 | 411 | 9.34 |
| **Parole Preparation** | | | | |
| psych_assess (0) | Manual | 231 | 99 | 42.9 |
| psych_assess (1) | | 321 | 55 | 17.1 |
| psych_assess (2) | | 87 | 5 | 5.75 |
| job_offer (0) | NLP/ tabular | 20185 | 4272 | 21.2 |
| job_offer (1) | | 14808 | 4218 | 28.5 |
| **Special Designation** | | | | |
| youth_offender (0) | Manual | 551 | 131 | 23.8 |
| youth_offender (1) | | 109 | 32 | 29.4 |
| elderly_parole (0) | Manual | 618 | 146 | 23.6 |
| elderly_parole (1) | | 67 | 22 | 32.8 |

### C. SOURCE AND DESCRIPTIVE STATISTICS FOR CONTINUOUS VARIABLES

| Variable | Source | Total n | Mean | Median | Q1 and Q3 |
|---|---|---|---|---|---|
| commissioner_rate (at denials) | NLP/ tabular | 26379 | 22.9 | 25.3 | [13.7, 29.6] |
| commissioner_rate (at grants) | NLP/ tabular | 8481 | 28.7 | 29 | [25.3, 33.4] |
| years_since_2007 (at denials) | NLP/ tabular | 26503 | 5.59 | 5 | [2, 9] |
| years_since_2007 (at grants) | NLP/ tabular | 8490 | 7.64 | 8 | [5, 11] |
| years_since_eligible (at denials) | NLP/ tabular | 22052 | 10.1 | 9 | [3, 16] |
| years_since_eligible (at grants) | NLP/ tabular | 7164 | 11.4 | 11 | [5, 17] |
| count_115s (at denials) | Manual | 466 | 8.23 | 5 | [2, 10] |
| count_115s (at grants) | Manual | 142 | 6.46 | 4 | [2, 7] |
| clean_time (at denials) | NLP/ tabular | 26503 | 7.19 | 5 | [2, 10] |
| clean_time (at grants) | NLP/ tabular | 8490 | 10.1 | 8 | [4, 14] |

## D. ROBUSTNESS CHECK USING FIXED EFFECT ON INDIVIDUAL COMMISSIONER

Table D: Rather than use `commissioner rate` to measure presiding commissioners, we use a fixed effect on the individual commissioners.

| **Data Source** | (a) Manual | (b) Tabular | (c) NLP |
|---|---|---|---|
| $n$ (Number of Hearings) | 688 | 34,993 | |
| **Hearing** | Adjusted Odds Ratio $e^{\beta}$ ($p$) | | |
| retained_attorney | **2.14 (0.01)** | **2.46 (0.00)** | **2.06 (0.00)** |
| initial_hearing | **0.44 (0.02)** | **0.40 (0.00)** | **0.47 (0.00)** |
| years_since_2007 | **1.20 (0.00)** | **1.14 (0.00)** | **1.22 (0.00)** |
| ethnicity_black | 0.87 (0.65) | 0.96 (0.24) | 0.95 (0.22) |
| ethnicity_latinx | 0.78 (0.44) | **1.12 (0.00)** | **0.89 (0.01)** |
| ethnicity_other | 0.77 (0.50) | **1.23 (0.00)** | 0.99 (0.87) |
| gender_female | 1.10 (0.86) | **1.24 (0.00)** | **1.30 (0.00)** |
| prison_is_level_iv | 1.02 (0.96) | **0.31 (0.00)** | **0.55 (0.00)** |
| offense_murder_first | 0.71 (0.28) | | 1.02 (0.56) |
| offense_murder_second | 0.70 (0.23) | - | **1.12 (0.00)** |
| offense_murder_attempt | 0.51 (0.11) | - | 1.11 (0.05) |
| years_since_eligible | 1.00 (0.87) | - | 1.00 (0.05) |
| precommit_gang | 0.99 (0.98) | - | **1.25 (0.00)** |
| tabe_edu_score | 1.27 (0.12) | - | **1.17 (0.00)** |
| psych_assess | **0.40 (0.00)** | - | **0.50 (0.00)** |
| clean_time | **1.08 (0.00)** | - | **1.02 (0.00)** |
| job_offer | **1.98 (0.02)** | - | **1.38 (0.00)** |
| programming_gang | - | - | **1.39 (0.00)** |
| programming_all | 1.09 (0.78) | - | - |
| 12steps_program_failed | **0.32 (0.04)** | - | - |
| victim_oppose | 0.51 (0.25) | - | - |
| victim_present | 0.39 (0.09) | - | **0.43 (0.00)** |
| district_attny_oppose | **0.21 (0.00)** | - | - |
| district_attny_present | | - | **0.68 (0.00)** |
| youth_offender | 0.95 (0.89) | - | - |
| elderly_parole | 0.91 (0.82) | - | - |
| crime_gang | 1.56 (0.34) | - | - |
| crime_drugs_alcohol | 0.61 (0.10) | - | - |
| precommit_sex_abuse | 0.73 (0.44) | - | - |
| justice_involved | 1.67 (0.05) | - | - |
| num_pris_convict_buc | 0.90 (0.84) | - | - |
| mental_illness | 0.92 (0.74) | - | - |
| mental_treatment | 1.25 (0.51) | - | - |
| count_115s | 1.00 (0.71) | - | - |
| chronos_bucket | **1.69 (0.02)** | - | - |
| attorney_opinion | 0.80 (0.50) | - | - |
| claim_innocence | 1.11 (0.81) | - | - |

| **Data Source** | (a) Manual | (b) Tabular | (c) NLP |
|---|---|---|---|
| **Fixed Effect** | Adjusted Odds Ratio $e^{\beta}$ ($p$) | | |
| anonymous_commissioner | **0.13 (0.02)** | **0.71 (0.00)** | **0.59 (0.00)** |
| anonymous_commissioner | 0.32 (0.53) | 1.36 (0.14) | 1.31 (0.24) |
| anonymous_commissioner | **0.04 (0.02)** | **0.24 (0.00)** | **0.17 (0.00)** |
| anonymous_commissioner | **0.04 (0.03)** | **0.25 (0.00)** | **0.24 (0.00)** |
| anonymous_commissioner | **0.22 (0.01)** | **0.68 (0.00)** | **0.59 (0.00)** |
| anonymous_commissioner | **0.13 (0.00)** | **0.79 (0.01)** | **0.74 (0.00)** |
| anonymous_commissioner | **0.17 (0.01)** | **0.70 (0.00)** | **0.65 (0.00)** |
| anonymous_commissioner | 0.19 (0.14) | 1.03 (0.85) | 0.95 (0.77) |
| anonymous_commissioner | 0.00 (1.00) | **0.15 (0.00)** | **0.08 (0.00)** |
| anonymous_commissioner | 0.55 (0.54) | **2.00 (0.00)** | **1.99 (0.00)** |
| anonymous_commissioner | 0.23 (0.21) | 1.04 (0.74) | 0.87 (0.32) |
| anonymous_commissioner | 0.36 (0.49) | **0.59 (0.01)** | **0.49 (0.00)** |
| anonymous_commissioner | **0.05 (0.02)** | **0.36 (0.00)** | **0.35 (0.00)** |
| anonymous_commissioner | 0.48 (0.42) | **0.76 (0.02)** | **0.70 (0.01)** |
| anonymous_commissioner | 0.00 (1.00) | **0.14 (0.00)** | **0.12 (0.00)** |
| anonymous_commissioner | 0.00 (0.99) | **0.47 (0.00)** | **0.39 (0.00)** |
| anonymous_commissioner | **0.02 (0.00)** | **0.62 (0.00)** | **0.50 (0.00)** |
| anonymous_commissioner | 0.00 (1.00) | 1.14 (0.55) | 1.16 (0.54) |
| anonymous_commissioner | **0.15 (0.00)** | **0.73 (0.00)** | **0.60 (0.00)** |
| anonymous_commissioner | **0.04 (0.01)** | **0.14 (0.00)** | **0.15 (0.00)** |
| anonymous_commissioner | 0.44 (0.43) | **0.60 (0.00)** | **0.46 (0.00)** |
| anonymous_commissioner | 0.00 (0.99) | **0.10 (0.00)** | **0.10 (0.00)** |
| anonymous_commissioner | **0.12 (0.00)** | **0.65 (0.00)** | **0.55 (0.00)** |
| anonymous_commissioner | **0.01 (0.00)** | **0.38 (0.00)** | **0.27 (0.00)** |
| anonymous_commissioner | 0.00 (0.99) | **0.51 (0.00)** | **0.41 (0.00)** |
| anonymous_commissioner | 0.11 (0.11) | **0.16 (0.00)** | **0.17 (0.00)** |
| anonymous_commissioner | **0.18 (0.01)** | **0.59 (0.00)** | **0.47 (0.00)** |
| anonymous_commissioner | 0.00 (0.99) | **0.51 (0.00)** | **0.48 (0.00)** |
| anonymous_commissioner | **0.07 (0.02)** | **0.67 (0.00)** | **0.57 (0.00)** |
| anonymous_commissioner | **0.05 (0.00)** | **0.59 (0.00)** | **0.47 (0.00)** |
| anonymous_commissioner | **0.04 (0.00)** | **0.49 (0.00)** | **0.42 (0.00)** |
| anonymous_commissioner | 0.09 (0.09) | 0.85 (0.28) | 0.81 (0.22) |
| anonymous_commissioner | **0.11 (0.01)** | **0.71 (0.00)** | **0.66 (0.00)** |
| anonymous_commissioner | **0.18 (0.01)** | 0.88 (0.16) | **0.73 (0.00)** |
| anonymous_commissioner | 0.00 (1.00) | **0.44 (0.01)** | 0.50 (0.06) |
| anonymous_commissioner | **0.06 (0.00)** | **0.47 (0.00)** | **0.38 (0.00)** |
| anonymous_commissioner | 1.50 (0.76) | 0.77 (0.16) | 0.73 (0.13) |
| anonymous_commissioner | 0.12 (0.06) | 0.94 (0.63) | **0.71 (0.03)** |
| anonymous_commissioner | 0.00 (0.99) | **0.55 (0.00)** | **0.45 (0.00)** |
| anonymous_commissioner | 0.17 (0.22) | **0.41 (0.00)** | **0.29 (0.00)** |
| anonymous_commissioner | **0.05 (0.02)** | **0.65 (0.00)** | **0.50 (0.00)** |
| anonymous_commissioner | 0.00 (0.99) | **0.25 (0.00)** | **0.20 (0.00)** |
| anonymous_commissioner | 0.00 (0.99) | **0.48 (0.00)** | **0.37 (0.00)** |
| anonymous_commissioner | 0.00 (0.99) | **0.14 (0.00)** | **0.13 (0.00)** |
| anonymous_commissioner | 0.00 (1.00) | 0.81 (0.30) | 0.66 (0.06) |
| anonymous_commissioner | **0.23 (0.02)** | 1.04 (0.63) | 0.96 (0.69) |
| anonymous_commissioner | 0.00 (1.00) | 0.73 (0.18) | 0.62 (0.07) |
| anonymous_commissioner | 0.00 (1.00) | **0.33 (0.00)** | **0.28 (0.00)** |
| anonymous_commissioner | 0.29 (0.27) | **0.57 (0.00)** | **0.49 (0.00)** |
| anonymous_commissioner | **0.12 (0.04)** | **0.70 (0.00)** | **0.56 (0.00)** |
| anonymous_commissioner | - | 0.00 (0.94) | 0.00 (0.95) |
| anonymous_commissioner | - | 0.00 (0.97) | 0.00 (0.96) |
| anonymous_commissioner | - | 0.00 (0.98) | 0.00 (0.98) |
| anonymous_commissioner | - | 0.64 (0.34) | 0.52 (0.21) |
| anonymous_commissioner | - | **0.12 (0.04)** | 0.00 (0.92) |
| anonymous_commissioner | - | 0.00 (0.98) | 0.00 (0.98) |
| anonymous_commissioner | - | 0.00 (0.98) | 0.00 (0.98) |
| anonymous_commissioner | - | 0.00 (0.98) | 0.00 (0.98) |
| anonymous_commissioner | - | 0.00 (0.95) | 0.00 (0.95) |
| anonymous_commissioner | - | 0.00 (0.93) | 0.00 (0.95) |
| anonymous_commissioner | - | 0.62 (0.28) | 1.53 (0.39) |

## E. REGRESSION USING ATTORNEY TYPE AS OUTCOME VARIABLE

Table E: Regressions onto attorney representation based on the subset of factors that are reasonably known to the candidate at the time they decide whether to retain an attorney, over the set of hearings where attorney representation status is known.

| **Data Source** | (a) Manual | (b) Tabular | (c) NLP |
|---|---|---|---|
| $n$ (Number of Transcripts) | 688 | 32,349 | |
| | Adjusted Odds Ratio $e^{\beta}$ $(p)$ | | |
| initial_hearing | 2.35 (0.26) | **0.91 (0.01)** | **0.90 (0.01)** |
| prison_is_level_iv | 0.21 (0.08) | **0.89 (0.01)** | 0.98 (0.62) |
| years_since_2007 | 0.94 (0.37) | **0.97 (0.00)** | **0.97 (0.00)** |
| ethnicity_black | **0.17 (0.00)** | **0.61 (0.00)** | **0.59 (0.00)** |
| ethnicity_latinx | **0.24 (0.01)** | **0.65 (0.00)** | **0.55 (0.00)** |
| ethnicity_other | **0.29 (0.04)** | **0.90 (0.02)** | **0.82 (0.00)** |
| gender_female | 2.68 (0.25) | **2.86 (0.00)** | **3.40 (0.00)** |
| offense_murder_first | 0.65 (0.44) | - | **1.19 (0.00)** |
| offense_murder_second | 1.36 (0.56) | - | **1.09 (0.00)** |
| offense_murder_attempt | 1.19 (0.79) | - | 0.97 (0.60) |
| years_since_eligible | 1.04 (0.18) | - | 1.00 (0.80) |
| precommit_gang | 1.10 (0.86) | - | **1.15 (0.00)** |
| precommit_sexual_abuse | 0.78 (0.65) | - | - |
| tabe_edu_score | 1.03 (0.90) | - | **1.22 (0.00)** |
| clean_time | 1.04 (0.21) | - | 1.00 (0.95) |
| job_offer | 2.18 (0.08) | - | **2.07 (0.00)** |
| programming_gang | - | - | **1.34 (0.00)** |
| programming_all | 0.71 (0.48) | - | |
| youth_offender | 2.52 (0.12) | - | |
| elderly_parole | 0.32 (0.13) | - | - |
| crime_gang | 1.42 (0.58) | - | - |
| crime_drugs_alcohol | **0.25 (0.00)** | - | - |
| claim_innocence | 0.69 (0.59) | - | - |
| justice_involved | 1.38 (0.44) | - | - |
| num_pris_convict_buc | 0.31 (0.22) | - | - |
| mental_illness | 1.40 (0.37) | - | - |
| mental_treatment | 1.50 (0.941) | - | - |
| count_115s | 1.02 (0.33) | - | - |
| chronos_bucket | **2.63 (0.01)** | - | - |

## F. DESCRIPTION OF METHOD USED TO CALCULATE WORD POLARITY SCORES

We model word usage as follows: $y \sim \text{Multinomial}\,(n, \pi)$. In our model, y is the vector of term word counts for the entire corpus, n is the total number of words in the corpus, and $\pi$ is the vector of probabilities for each word in the vocabulary. To account for inherent differences in word usage not based on the examined feature, the model is typically instantiated with a Dirichlet prior with parameter vector $\alpha$, a vector of counts for each word in the corpus. Intuitively, $\alpha$ can be thought of as the number of times each word has been encountered before examining the corpus. For our experiments, we set $\alpha$ to be the vector of word counts across all attorney speech in all hearings, regardless of the attorney's status.

Given an observed vector of word counts from the corpus, *y*, the prior distribution, and the total number of words in the corpus *n*, the maximum likelihood estimate of the underlying probability distribution

$$\pi \text{ } is \text{ } \hat{\pi} = \frac{1}{n+\alpha_0} \cdot (y + \alpha)$$

where $\alpha_0$ is the sum of $\alpha_w$ for each word *w* in the corpus. We let *a* and *b* indicate the disjoint subsets of our corpus yielded by the feature under examination using superscripts, such that $y^{(a)}$ indicates the vector of word counts for that particular subset, with $\alpha^{(a)}$ and $n^{(a)}$ defined analogously. Under these specifications, we can estimate the odds of a specific word *w* compared to others for a subset *a* as:

$$\hat{\Omega}_w^{(a)} = \frac{\hat{\pi}_w^{(a)}}{1-\hat{\pi}_w^{(a)}}$$

We then estimate the log-odds ratio for the word *w* between the two groups *a* and *b* (denoted $\hat{\delta}_w^{(a-b)}$) as follows:

$$\hat{\delta}_w^{(a-b)} = \log \frac{\left(y_w^{(a)} + \alpha_w^{(a)}\right)}{n^{(a)} + \alpha_0^{(a)} - y_w^{(a)} - \alpha_w^{(a)}} - log \frac{\left(y_w^{(b)} + \alpha_w^{(b)}\right)}{n^{(b)} + \alpha_0^{(b)} - y_w^{(b)} - \alpha_w^{(b)}}$$

One of the important benefits of using a model-based approach (as opposed to just computing the log odds ratio directly from the vector of word counts), is that it offers not just a score for each word, but an estimate of the variance for that score. In particular, the variance is estimated as:

$$\hat{\delta}^2(\hat{\delta}_w^{(a-b)}) = \frac{1}{(y_w^{(a)} + \alpha_w^{(a)})} + \frac{1}{(y_w^{(b)} + \alpha_w^{(b)})}$$

So, instead of reporting the raw scores for any given word, we can instead report the normalized z-score, defined as:

$$z_w^{(a-b)} = \hat{\delta}_w^{(a-b)} \Big/ \sqrt{\hat{\sigma}^2 \left(\hat{\delta}_w^{(a-b)}\right)}$$