



Three Essays on Rule Enforcement

Citation

Sigstad, Henrik. 2019. Three Essays on Rule Enforcement. Doctoral dissertation, Harvard University, Graduate School of Arts & Sciences.

Link

<http://nrs.harvard.edu/urn-3:HUL.InstRepos:42013070>

Terms of use

This article was downloaded from Harvard University's DASH repository, and is made available under the terms and conditions applicable to Other Posted Material (LAA), as set forth at

<https://harvardwiki.atlassian.net/wiki/external/NGY5NDE4ZjgzNTc5NDQzMGIzZWZhMGFIOWI2M2EwYTg>

Accessibility

<https://accessibility.huit.harvard.edu/digital-accessibility-policy>

Share Your Story

The Harvard community has made this article openly available.
Please share how this access benefits you. [Submit a story](#)

Three Essays on Rule Enforcement

A dissertation presented

by

Henrik Sigstad

to

The Department of Economics

in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

in the subject of

Economics

Harvard University

Cambridge, Massachusetts

July 2019

© 2019 Henrik Sigstad

All rights reserved.

Dissertation Advisors:
Professor Alberto Alesina
Professor Edward Glaeser

Author:
Henrik Sigstad

Three Essays on Rule Enforcement

Abstract

When can individuals be incentivized to enforce impersonal rules? This dissertation consists of three chapters, each trying in different ways to address this question. The first chapter uses a close election regression discontinuity design to document that local politicians in Brazil are more likely to be acquitted from corruption charges if they win the election. I provide evidence consistent with the effect coming from law enforcers being directly influenced by the power of elected politicians, and not from electoral winners having better lawyers. The second chapter exploits random assignment of court cases to judges to ask to what extent the identity of the judge matters for whether criminal cases are appealed in Norway. I find that there are large differences between the rate at which different judges have their decisions appealed, which cannot be attributed to statistical noise. Judges with higher appeal tendencies are found to be more stringent, faster, disagreeing more with their peers, and having spent a longer time in their current court. The third chapter discusses theoretically when norms of collective action can be enforced and argues that they are more likely to be enforced if signals about norm violations are public and categorical in nature.

Contents

Abstract	iii
Acknowledgments	ix
Introduction	1
1 Judicial Subversion: Evidence from Brazil	3
1.1 Introduction	3
1.2 Institutional context	9
1.2.1 The legal remedies against political corruption in Brazil	9
1.2.2 Brazilian trial court judges and prosecutors	11
1.2.3 Brazilian local government	12
1.3 Data	13
1.3.1 Judicial data	13
1.3.2 Electoral data and matching	15
1.3.3 Summary statistics	18
1.4 Empirical strategy	18
1.5 Main results	21
1.6 Mechanisms	25
1.6.1 Direct tests of the lawyer hypothesis	28
1.6.2 The location of the court	31
1.6.3 Cases decided early versus late in the electoral term	33
1.6.4 Do electoral winners have more at stake?	34
1.6.5 Liquidity constraints	35
1.6.6 Which law enforcers are influenced?	36
1.6.7 Other mechanisms	37
1.7 Incentives for corrupt politicians to run for election	38
1.8 Implications and future research	40
2 Judges, Appeals and Judicial Decision Making	41
2.1 Introduction	41
2.2 Institutional Setting	45

2.2.1	The Norwegian Court System	45
2.2.2	Appeals of District Court Decisions to the Courts of Appeal	48
2.3	Data Sources and Research Design	50
2.3.1	Data and Sample Selection	50
2.3.2	Testing Random Assignment of Cases to Judges	52
2.3.3	Regression Model	55
2.3.4	Judge Stringency and Judge Appeal Tendency	57
2.3.5	Other Judge Characteristics	60
2.4	Judge Appeal Tendency and Appeal Behavior	62
2.4.1	Main Results: The Effects of Judge Appeal Tendency on Appeal Outcomes	62
2.4.2	The Role of Judge Stringency	67
2.4.3	The Role of Other Judge Characteristics	70
2.4.4	Judge Appeal Tendencies and Career Outcomes	74
2.5	Judge Appeal Tendency, Judicial Costs and Defendant Outcomes	76
2.5.1	The Effects of Judge Appeal Tendency on Judicial Costs	76
2.5.2	The Share of Appeals Due to Judges with High Appeal Tendencies	78
2.5.3	The Effects of Judge Appeal Tendency on Defendants' Future Outcomes	79
2.6	Conclusion	81
3	Coordinated Enforcement	82
3.1	Introduction	82
3.2	Relation to Prior Literatures	86
3.3	A Model of Coordinated Enforcement	87
3.3.1	Characterizing Which Signals Can Trigger Sanctions	87
3.3.2	The Connection With Higher Order Beliefs	89
3.3.3	The Connection With Deterrence	90
3.4	Special Cases	92
3.4.1	Categorical Norms	92
3.4.2	The Role of Observability and Correlation	94
3.5	Discussion	96
3.5.1	Comparative Statics	97
3.5.2	Welfare Implications	97
3.5.3	Prescriptions	98
3.5.4	Implications for Crime and Punishment	99
3.5.5	Characterizing Pro-social Preferences	99
3.5.6	Intuition for Repeated Games and Future Works	100
	References	102

Appendix A	Appendix to Chapter 1	108
A.1	Parsing court documents	108
A.2	Incumbency disadvantage	109
Appendix B	Appendix to Chapter 2	112
B.1	Interpretation of the First-Stage Coefficient in Random Judge Designs	112
B.2	Proofs for Propositions 1–3	114
B.3	Monte Carlo Simulations	117
B.4	Measuring Judge Disagreement	119
Appendix C	Appendix to Chapter 3	121
C.1	Characterizing Which Signals Can Trigger Sanctions	121
C.2	The Connection With Higher Order Beliefs	122
C.3	The Connection With Deterrence	122
C.4	Categorical Norms: An Example With Uniform Distributions of States and Signals	123
C.5	Categorical Norms: A Result for Arbitrary Distributions of States and Signals	124
C.6	The Role of Observability and Correlation	124

List of Tables

1.1	Summary statistics	17
1.2	Balance on pre-election variables	20
1.3	Main regression discontinuity results	22
1.4	The effect of winning the election on the quantity and quality of lawyers . .	29
1.5	Heterogeneous effects	32
1.6	The filing of corruption cases and future elections	39
2.1	Testing for Random Assignment of Criminal Cases to Judges.	54
2.2	The Effects of Judge Appeal Tendency on Appeal Outcomes.	63
2.3	Decomposition of the Variance of Raw Judge Appeal Tendency.	66
2.4	The Effects of Judge Stringency on Appeal Outcomes.	68
2.5	Decomposition of the Variance of Systematic Judge Appeal Tendency.	70
2.6	The Effects of Other Judge Characteristics on Appeal Outcomes.	72
2.7	Judge Appeal Tendency, Judicial Tenure and Career Progression.	75
2.8	The Effects of Judge Appeal Tendency on Judicial Costs.	78
2.9	The Effects of Judge Appeal Tendency on Defendant's Recidivism and Future Employment.	80
A.1	Incumbency disadvantage	111

List of Figures

1.1	Coverage of judicial data by state	14
1.2	The number of Ações de Improbidade by state	16
1.3	Regression discontinuity plot	23
1.4	Regression discontinuity estimates	24
1.5	Placebo regression discontinuity plot	25
1.6	McCrary sorting test	26
1.7	The number of new lawyers over the electoral cycle	30
2.1	Case Trial and Appeal Process in Norway's Criminal Justice System.	49
2.2	The Distribution of Judge Appeal Tendency.	56
2.3	The Distribution of Leave-Out and Systematic Judge Appeal Tendency.	64
3.1	The Coordination Game, G	88
B.1	The Empirical Distribution of t -values from Monte Carlo Simulations.	118

Acknowledgments

I am grateful to my advisers Alberto Alesina, Edward Glaeser and Melissa Dell for their excellent guidance. For the work on the first chapter I would like to give a special thanks to Luis Fabiano de Assis and Bruno Bodart for teaching me about the workings of the Brazilian legal system, and for organizing meetings with lawyers, judges and prosecutors throughout Brazil. Thanks to Saulo Benchimol for research assistance. Also thanks to Rolf Aaberge, Daron Acemoglu, André Assumpção, Priscila Beltrami, Fernando Bizarro, Taylor Boas, Washington Brito, Filipe Campante, Fabrício Fernandes de Castro, Moya Chin, Paulo Costa, João Falcão, Eliana la Ferrara, Leopoldo Ferguson, Isabela Ferrari, João Victor Freitas Ferreira, Ray Fisman, Felipe Fontes, Siddharth George, Estêvão Gomes, Daniel Hidalgo, Moshe Hoffman, Liana Issa, Louis Kaplow, Asim Khwaja, Alejandro Lagomarsino, Leany Lemos, Horacio Larreguy, Victor Leahy, Jetson Leder-Luis, Debora Maliki, Etienne Martins, Nathan Nunn, Tzachi Raz, Mark Ramseyer, Ivan Ribeiro, Carlos Sanchez Martinez, Raul Sanchez de la Sierra, Jesse Shapiro, Andrei Shleifer, Holger Spamann, Matt Stephenson, Anderson Summa, Rafael Di Tella, Pedro Tepedino, Clémence Tricaud, Erez Yoeli, Luciana Zaffalon, and seminar participants at Harvard, MIT, University of Oslo, Universidade de Brasília, Society for Institutional and Organizational Economics, Ridge-LACEA political economy, Tribunal Regional Federal da 2a Região, and the 6th Global Conference on Transparency Research. I acknowledge the support of the Haddad Foundation, through the Harvard-Brazil Cities Research Grant Program of the David Rockefeller Center for Latin American Studies, and the Lynde and Harry Bradley Foundation. For the work on the second chapter I am grateful to Baard Marstrand at the Norwegian Courts Administration (Domstoladministrasjonen) for help accessing the data and understanding institutional details, and Martin Eiebakke and Eirik Dyrstad in the Norwegian Courts Commission (Domstolkommisjonen). I acknowledge generous financial support from the Norwegian Research Council, through the project 240653. For the work on the third chapter I thank Andrew Ferdowsian for research assistance.

To Frida and Alfred

Introduction

The main motivation behind this dissertation is that I believe that societies governed by impersonal rules often are better societies. I believe that rules can lead to a predictable investment climate, reduce violence and inequalities, and help in the provision of public goods. However, rules in themselves are just words written on paper or something existing only in our minds. It has thus been a great puzzle to me, and still is, how to incentivize individuals to actually enforce the rules. While I am not able to answer this puzzle in the current dissertation, I believe that each of the three chapters are somehow trying to address it.

The first chapter asks whether a formally independent judicial system is able to check abuses of political power. Together with Guilherme Lambais I estimate that candidates in Brazilian local elections who narrowly win an election are 60 percentage points more likely than candidates who narrowly lose to be acquitted in corruption cases filed against them before the election. We show evidence suggesting the effect comes from a direct influence of the power of elected politicians over law enforcers, rather than from elected politicians having better lawyers, and that it might lead to an adverse selection of corrupt politicians in electoral offices.

In the second chapter, written together with Manudeep Bhuller, I ask to what extent the identity of a judge matters for whether criminal cases are appealed in Norway, and what might be driving such differences across judges. Using random assignment of cases to judges in trial courts and a novel dataset linking trial court decisions to outcomes in appellate courts, we document a series of stylized facts about judicial decision making

and the appeal process. First, we show that there is striking variation in the rate at which decisions are appealed across trial court judges. While the average appeal rate is 15%, one out of ten judges have their decisions appealed at a rate of 21% or more. Importantly, we show that the leave-out judge appeal tendency for the assigned judge is highly predictive of appeals and decision reversals in appellate courts. Our findings indicate that 20% of the variance in appeal rates across judges can be attributed to *systematic* judge differences in appeal tendencies. Such differences remain after controlling for differences in judge tendencies to convict, incarcerate, and impose long sentences. We estimate that 4% of all appeals could be eliminated by replacing judges with higher than average appeal rates with average judges, conditional on judge stringency. Second, we find that court cases are more likely to be appealed if they are assigned to judges with lower average case processing time, to judges with deviant judicial preferences, and to judges who have spent a long time in the same court. Third, we find that judges with lower appeal rates are less likely to find a new job in another court, with the possible exception of larger courts with many applicants. Finally, we find no evidence that the appeal tendency of an assigned trial court judge causally affects defendants' recidivism or future labor market outcomes.

In the third chapter, written together with Moshe Hoffman, N. Aygun Dalkiran, and Erez Yoeli, I study theoretically how norms of collective action might be enforced. We argue that such norms are often maintained via a coordinated behavior among norm enforcers. Examples of this include when democratic norms are maintained via public protest, or when international norms are maintained via multilateral sanctions. We present a stylized model of such coordinated enforcement: enforcers receive signals indicative of the agent's action, and payoffs from sanctioning, according to a state-independent coordination game. We characterize which signals can trigger sanctions, what actions can be deterred, and the relation to higher order beliefs. Two special cases—a variant of global games, and another that considers signals' type II error and correlation—capture the categorical nature of norms and the undue influence of plausible deniability.

Chapter 1

Judicial Subversion: Evidence from Brazil¹

1.1 Introduction

There are many reasons to believe that keeping elected politicians accountable to the law is important. It can prevent outright stealing of public funds, and makes sure the rules governing democratic elections are abided by. Furthermore, it might play an important role in fostering economic development by creating a predictable investment climate shielded from arbitrary government expropriations (North and Weingast, 1989). Finally, it can reduce political violence as disputes with the government can be solved by legal means. Yet, such an outcome might be difficult to achieve in practice. Judges and other agents of justice can face strong incentives to not enforce the law in the face of powerful politicians. Elected officials might be able to make their careers terrible, by denying them promotions, removing them from office, or transferring them to disagreeable locations or positions. Sometimes politicians in power might even be allied with dangerous militias or criminal groups making it life-threatening go against them. Friendly law enforcers, on the other hand, might be rewarded with easier access to government jobs, contracts, or public services, to

¹Co-authored with Guilherme Lambais

themselves or to their friends and family. Finally, politicians might starve the budgets of law enforcement institutions should they turn disloyal, or judges might fear that any decision against powerful politicians will not be enforced, as the enforcement of their decisions is typically in the hands of the executive. This is not to mention that politicians also have an important role in *making* the laws, and might create legal loopholes that lets them get away with harmful activities even when judges do apply the letter of the law.

In response to this, most modern societies have imposed an array of rules to make the judicial system more immune to political influence. These include laws that secure the tenure of judges, making them virtually impossible to remove or transfer to a different position for politicians who dislike their rulings, and rules which prohibit judges from most outside jobs, especially jobs in the executive. The power to nominate and promote judges has also been taken from the hands of politicians and given to judicial councils, or the judiciary itself, in a majority of countries (Garoupa and Ginsburg, 2009). There are often similar rules protecting the independence of prosecutors. How well do these rules work? Are they sufficient to remove all influence of political power over judicial decisions, or is it still the case, even with all these safeguards, that elected politicians tend to face a more lenient justice? Apart from anecdotal evidence we know very little about this. There are some studies showing that elected politicians or their party are favored by the judiciary when they control the nomination or promotion of judges (Ramseyer and Rasmusen, 2001; Sanchez-Martinez, 2017; Poblete-Cazenave, 2019; Mehmood, 2019). However, as far as the authors are aware, we have no causal evidence showing whether politicians holding electoral offices are favored even in settings where they have no control over the judges' career path. The contribution of this paper is to provide such evidence.

To do this we study a type of corruption court cases called *Ações de Improbidade* involving local politicians in the trial courts of the Brazilian state judiciaries. Brazilian trial courts are ideal for studying this question since all the formal ways of ensuring that judges are immune to political influence are in place. They are very difficult to remove, cannot be transferred to other positions against their will, and cannot have other jobs except teaching.

Politicians have no influence over their careers: They are appointed by a competitive exam administered by the appeals court (Tribunal de Justiça) who also determines promotions. Finally, the judges receive a very high salary placing them among Brazil's top earners. Similar rules apply to the public prosecutors who are in charge of the prosecution.

Is this enough to prevent politicians in power from having influence over judicial outcomes? Answering this question is challenging. Showing that elected politicians tend to win at a higher rate in court than others does not prove that decisions are influenced by their power, it could just be that they tend to face more frivolous cases, perhaps due to politically motivated litigation. Conversely, no difference in win rates between elected politicians and other litigants does not prove that the judicial system is immune to political influence, it could just be that prosecutors are only filing cases against powerful politicians when they have exceptionally strong evidence. We solve this empirical challenge by using a close election regression discontinuity design, focusing on corruption cases which are filed before and decided after the election. If close elections are decided at random, marginal electoral winners and marginal electoral losers will on average have similar types of corruption cases pending against them. The strength of the evidence or the gravity of the misconduct should not differ systematically. Thus, if we find that marginal electoral winners are more likely to be acquitted than the marginal losers, we would thus be able to conclude that this difference is causally due to the election result, and not to anything related to the initial strength of the case.

In our main regression we find large effects of political power on judicial outcomes. The point estimate indicates that marginal electoral winners are 60 percentage points more likely to win in court than marginal electoral losers, a difference which is statistically significant at the one percent level. There is no effect of the election on corruption cases decided *before* the election, indicating that marginal winning and marginal losing candidates do tend to be involved in corruption cases with similar strength of evidence. The effect is larger if a politician is elected mayor, but is also present for candidates to the local legislature.

In the second part of our analysis we try to distinguish between two main types of

mechanisms that can explain the result. First, it could be due to the winners of the election investing more resources into winning the case by legal means, by hiring better lawyers. Second, it could be due to a *direct* influence of political power over law enforcers such as judges, prosecutors, judicial staff, or witnesses. This could be due to the promise of easier access to public services, jobs or contracts for themselves or their friends and families, threats by criminal groups allied to the elected politician, or just a purely psychological effect that power may inflict on the minds of law enforcers. We provide several pieces of evidence indicating that the result is unlikely to be only driven by the first channel. First, we directly measure the quantity and quality of the lawyers who are registered on each court case, and find no evidence that marginal electoral winners tend to increase the number or the quality of lawyers on their case after the election, compared to marginal electoral losers. We also find no tendency of neither the marginal winner nor the marginal loser of hiring new lawyers at a higher rate after the election than before. Thus, if electoral winners are more likely to win in court due to receiving superior legal counsel it has to be either by lawyers who are not registered on the case, or by the lawyers registered on the case exerting more effort. The second piece of evidence against the better lawyer hypothesis is that the effect is larger if the court is located in the municipality of the politician than if the court is located in a nearby municipality. If the effect is only driven by winners having better lawyers, we would predict no such a difference since the legal means with which a lawyer can win a case does not depend on the location of the court. On the other hand, the *non*-legal means that an elected politician can use to win a case, such as offering jobs and contracts to relatives of law enforcers, is likely to be much higher if the court is located in the municipality where the politician holds power. The third piece of evidence is that the effect of winning the election on judicial outcomes seems to be larger for cases decided early in the electoral term. This is consistent with the effect being driven by the non-legal channel since the future opportunities for rewarding friendly law enforcers is largest just after the politician takes office. On the other hand, if the effect is due to electoral winners having better lawyers we would expect the effect to be larger if the case is decided late in the term,

which would give the lawyers more time to have an impact.

Next, we argue that an important reason to believe that electoral winners might want to get better lawyers, that they might have more at stake in the case given that they might lose office, does not apply. Even though loss of office is a possible penalty in *Ações de Improbidade* the penalty does not come into effect as long as the politician has opportunities to appeal the decision. As the Brazilian judiciary is often slow, has effectively four instances, and a myriad of ways to appeal decisions, this means that no electoral winner in our sample has actually had to step down due to being convicted in an *Ação de Improbidade*. The risk of not being able to run for future electoral offices due to an *Ação de Improbidade* is, however, real. But this applies to electoral losers as well. In fact, since there is an incumbency *disadvantage* among Brazilian mayors, marginal losers might actually have more at stake in *Ações de Improbidade* than marginal winners. Finally, we show that there is no evidence that the effect is driven by another reason for marginal winners having better lawyers: That they are less liquidity constrained. The point estimate of the effect is identical for politicians with and without higher education, and larger for politicians with a larger than median campaign.

Our main result which shows that electoral winners are more likely to be acquitted from corruption charges than electoral losers, might lead to an adverse selection of politicians in electoral offices since politicians facing corruption charges could seek electoral office to protect themselves from the law. In order to investigate whether such an effect might be going on, we run a panel regression including all politicians who has ever run for the office of mayor in the elections between 2000 and 2016. We find that a politician is 8 percentage points more likely to enter a mayoral race when an *Ação de Improbidade* is filed against them in the four years leading up to the election, than what would otherwise have been predicted by her past electoral career. This is true controlling for politician fixed effects, election year dummies, and quadratic politician trends. While we cannot rule out other explanations for this result, it is at least consistent with the view that corrupt politicians could be induced to run for electoral office in order to receive a more favorable treatment by

the law.

The question of creating a judicial system that is immune to the subversion of the politically powerful has captivated researchers at least since Montesquieu (1989). Approaches vary from theoretical (e.g. Glaeser and Shleifer, 2002), to case studies (e.g. Chavez, 2004), and cross-country regressions (e.g. La Porta *et al.*, 2004). Yet, there is little within country quantitative analysis. Two notable exceptions are Ramseyer and Rasmusen (2001) who show that Japanese lower court judges who are lenient in cases involving the government tend to be promoted at a higher rate, and Helmke (2005) which shows that Argentinian judges started to rule against the government when it became clear that the ruling party would lose in the coming elections. For studies which claim to offer *causal* evidence that elected politicians are favored by courts, the authors are only aware of Sanchez-Martinez (2017) and Poblete-Cazenave (2019), both making use of close election regression discontinuity designs. Our paper differ from these along several dimensions. Most importantly, both take place in settings where politicians have power over the careers of law enforcers, whereas in our setting local politician have no control over the careers of neither judges, prosecutors, nor judicial staff. We also differ from Sanchez-Martinez (2017) by looking at court cases involving politicians themselves, rather than members of the winning political party. By showing that judges are biased towards elected politicians we contribute to the large literature on judicial bias, which have tended to focus on ethnic biases (e.g. Shayo and Zussman, 2011; Alesina and Ferrara, 2014; Arnold *et al.*, 2018; Abrams *et al.*, 2012). A key challenge faced by this literature it that it is typically not possible to know whether the bias is *due* to the ethnicity of the litigant or to something correlated with ethnicity. We contribute by showing that judges are biased towards the politically powerful, and that this bias is *causally* due to gaining political power and not to something correlated with being powerful. Finally, we contribute to the broader literature on how to prevent political corruption. This literature has mostly focused on when corruption can be disincentivized by voters (see Olken and Pande, 2012) or auditors (e.g. Avis *et al.*, 2018; Ferraz and Finan, 2008, 2011), we contribute by adding evidence about when corruption can be disincentivized by judges.

The rest of the paper is organized as follows. Section 1.2 describes the legal remedies against corruption in Brazil, the careers of Brazilian judges and prosecutors, and which tools Brazilian local politicians have at their disposal to benefit or harm law enforcers. In Section 1.3 we discuss how we constructed a data set of corruption cases involving local politicians. Section 1.4 presents the empirical strategy we use to estimate the effect of being elected on judicial decisions. In Section 1.5 we present our main results showing that politicians are more likely to be acquitted of corruption charges if they win the election. Section 1.6 discusses to what extent the result is driven by electoral winners having better lawyers, or by a direct influence of political power. In Section 1.7 we show that politician have a higher probability of running for electoral office after an *Ação de Improbidade* is filed against them. Section 1.8 concludes.

1.2 Institutional context

In this section, we first describe the legal remedies against corruption among elected officials in Brazil, with a focus on *Ações de Improbidade*. We then describe the judges and the prosecutors who are involved in the cases. Finally, we describe the relevant features of Brazilian local government.

1.2.1 The legal remedies against political corruption in Brazil

There are several legal remedies against corruption among elected officials available in Brazil. First, corruption is defined as a crime in the penal code, and there are various other corruption related crimes such as money laundering. Then there are three types of civil suits which can be brought against corruption and less serious administrative malfeasance: *Ação Civil de Improbidade Administrativa* ("Ação de Improbidade"), *Ação Civil Pública*, and *Ação Popular*. In this paper we will not look at the criminal cases against corruption. Criminal cases against mayoral candidates are sent to the appeals court in the case they win the election (a rule colloquially known as *foro privilegiado*), making it difficult to interpret a close election regression discontinuity estimate for these cases. Among the civil suits, we will

focus on Ações de Improbidade for two reasons. First, they are the most serious of the civil suits.² Second, the other civil suits can be settled which makes it difficult to interpret regression discontinuity estimates.

Ações de Improbidade can be filed against any act by a public official which causes either violation of administrative principles, damage to the treasury or illicit enrichment. Typical cases involve the hiring of public workers without proper procedure and fraud in government contracting. The suit can only be brought by the public prosecutor or the entity harmed by the corrupt act. For cases involving local politicians, the latter is typically the municipality. The possible penalties are loss of office, loss of political rights for 3-10 years, reimbursing the treasury, fines up to 100 times the monthly wage, and the prohibition of receiving government contracts for 3-10 years. Loss of political rights is seen as one of the most severe penalties, since it includes not being able to run for electoral office. The judge has a wide discretion in deciding which penalties to apply. One feature of these cases which makes interpreting our results easier is that they cannot be solved by a settlement between the parties. The cases filed by the public prosecutor are typically initiated by someone filing a complaint to the prosecutor, then the prosecutor investigates (*inquérito civil*) and chooses whether to file a case depending on the outcome of the investigation. We will only look at the cases which are filed by the public prosecutor in this paper, not those filed by the municipality: The mayor is the legal representative of the municipality, thus if a mayoral candidate who is facing a case filed by the municipality wins the election she will end up becoming both the plaintiff and the defendant, and the case is often dismissed by the judge.

An example of a typical case in our data is an Ação de Improbidade filed by the public prosecutor against the mayor of the municipality Fartura in the interior of São Paulo. The mayor had awarded a contract to provide fuel to the municipality to a firm owned by the son of the vice-mayor, and it is illegal to contract a firm which is under the influence of a public servant. In the decision the judge agreed the contract was illegal, ruled it void, and

²Ação Popular can only reverse political or administrative decisions, and leads to no further penalties for the politician, whereas Ação Civil Pública can only lead to fines and injunctions.

imposed small fine on the vice-mayor. However, the judge acquitted the mayor arguing the misconduct was not done in "bad faith", and did not impose any repayment of funds arguing it was not proven that the contract had lead to a financial loss to the municipality.³

1.2.2 Brazilian trial court judges and prosecutors

Judges at the trial courts in Brazil's state judiciaries are formally very independent of politics. They get appointed via a competitive exam administered by the appeals court, receive tenure after two years of service, and are promoted by the appeals court on criteria which alternate between seniority and merit. The only formal influence of politics is that one fifth of the appeals court judges are nominated by the state governor.⁴ While judicial corruption is recognized as a problem, there is a general belief that trial court judges are not easily corruptible. The main reason cited is that judges earn very high wages: Just the official wage of a trial judge puts her among Brazil's top earners, not counting several perquisites such as housing allowances.

The public prosecution (*Ministério Público*) in Brazil is also formally very independent from both the executive and the judiciary, and is often called the fourth branch of government. Public prosecutors are appointed via a competitive exam administered by the state chief prosecutor, receive tenure after two years of service, and are promoted by the chief prosecutor. The chief prosecutor is, however, appointed by the governor for a two year term with the possibility a one term renewal. In practice, the governor chooses the chief prosecutor from a list of three candidates prepared after a vote among the prosecutors. Prosecutors have wide discretion in deciding which cases to work on. But any decision to drop a case is subject to review by the chief prosecutor.

³An important feature of *Ações de Improbidade* is that it is typically not enough to demonstrate that the law was broken, but also that it was broken in "bad faith". This is a very common reason for acquitting politicians. Sentences including the words *não* ("no") and *dolo* ("intentional misconduct") appear in 70% of all acquittals for which we have the legal justification.

⁴The seats are filled alternately by public prosecutors and lawyers. The organizations representing public prosecutors or lawyers prepare a list of six candidates, which is reduced to three by the appeals court, before the governor chooses. Another potential source of judicial influence of the *governor* is the control over the budget of the state judiciary (see Zaffalon, 2018).

Each state is divided into judicial districts, which typically covers between one to three municipalities each. If a district is composed of several municipalities, the court is typically located in the largest municipality. A judicial district might have between one to several hundred judges, depending on its size. In districts with many judges, there are judges specialized in certain areas such as civil and criminal cases. Cases are randomly allocated to judges if there are more than one judge who has jurisdiction. The public prosecution has a parallel structure to the judiciary with sections in each judicial district.

1.2.3 Brazilian local government

Local government in Brazil is composed of the mayor (*prefeito*) and the city council (*câmara de vereadores*). Mayors are elected via a first past the post electoral system, with the exception of cities with a population greater than 200,000 which have a second round run-off between the two top candidates if none received more than 50% of the votes. The city council is filled by an open list proportional representation system. Each candidate is part of an electoral coalition composed of several parties. The coalitions receive seats in the council according to the vote share received by all candidates within the coalition, and the seats are allocated to the candidates who received most votes.

The responsibilities of the mayor is to administrate the city budget, in areas such as education, health and transport, and to collect municipal taxes. In doing this the mayor has the power to contract firms and hire municipal workers. Many municipal employees are hired via a competitive civil service exam and receive tenure after three years of service, but there are some categories of jobs which gives the mayor much more discretion in deciding who to hire, such as commissioned posts (*cargo em comissão*), positions of trust (*função de confiança*), and temporary jobs.⁵ Mayors often make use of these job categories to circumvent the civil service exam (see Colonnelli *et al.*, 2017; Akhtari *et al.*, 2018). In fact, a large share of

⁵The difference between commissioned posts and positions of trust is that commissioned posts can be given to any person that satisfy some rules if any (*e.g.* some posts may require a higher education diploma), while positions of trust can only be given to people who are already public servants and got the position via civil service exam (*servidor público concursado*).

the corruption cases in our sample are about the misuse of these categories to hire workers who should have been hired via the civil service exam. Finally, the mayor can propose municipal laws. The chief role of the city council is to approve municipal laws. However, city councilors also have the power to hire certain workers, including commissioned posts both for their office staff, such as policy and advisement jobs, and to run the city council in its various departments, such as communications, legal, and IT. In addition, the council members have to write up and vote on the annual budget for the municipality every year, which is approved into law (*Lei Orçamentária Anual*). In doing so, depending on the power of the council member, he can have some influence on how the budget will be spent, although its execution is delegated to the mayor. There are a myriad of ways in which mayors and city councilors could use public resources to affect the life of law enforcers. While judges and prosecutors personally are not allowed to be hired by the municipality or receive government contracts, there are no such prohibition for their relatives and friends. Also, judges and prosecutors, who are required to live in the judicial district, depend on municipal services such as water, electricity, and health care. It is not unlikely that elected local politicians would have it in their power to make access to such services easier for certain law enforcers.

1.3 Data

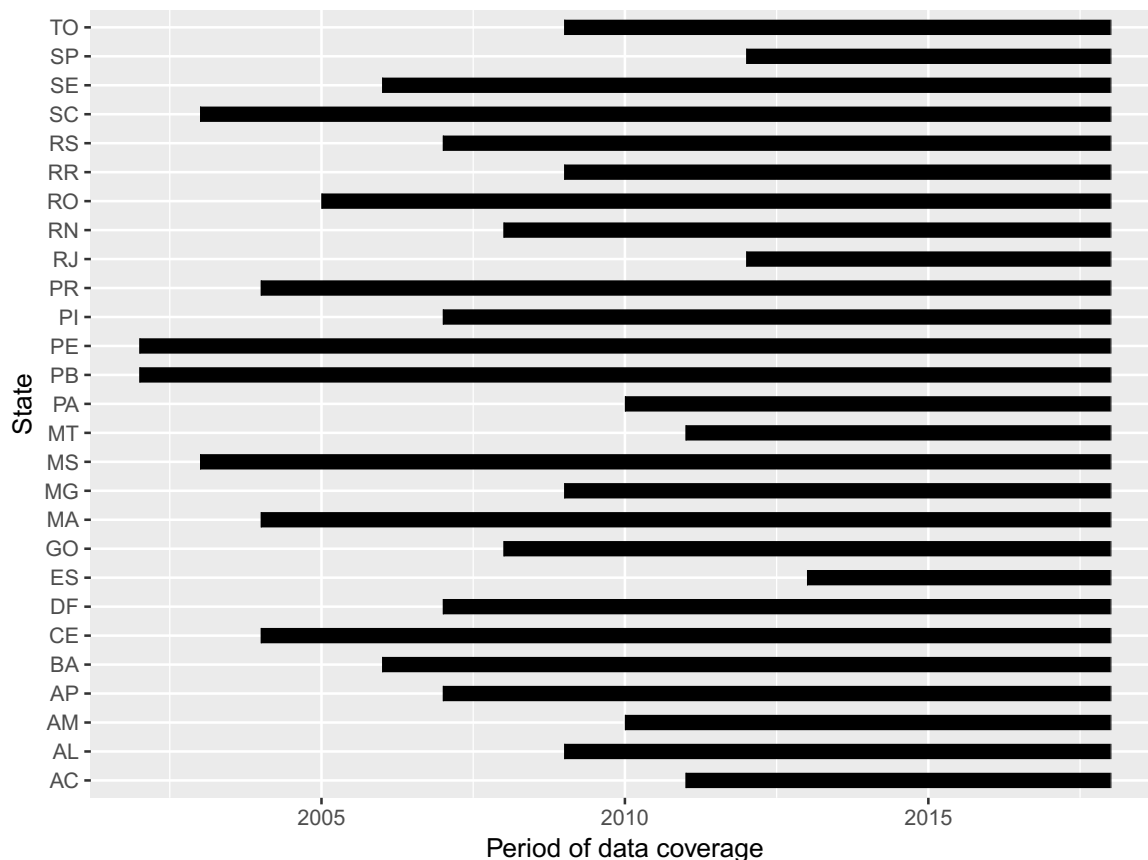
In this section we explain how we constructed our main data set of Ações de Improbidade involving local politicians, and present summary statistics.

1.3.1 Judicial data

For all states except São Paulo our data comes from the daily official publications of the courts called *Diários de Justiça*. The courts are required by law to publish all judicial decisions in these outlets, and they also publish several other minor statements about the case as it proceeds. This means thousands of pages of raw text every day. In order to generate a data set from such large source of raw text we employ a technique called *parsing* which is a method commonly used in computer science to extract the meaning of computer languages,

and in computational linguistics to extract the meaning of natural language sentences. Parsing is a useful method for machine reading court documents, since such documents are typically very structured. The parser is being told the logical structure of the text, and use this to generate a data set with information about the cases such as the type of lawsuit, the names of litigants, and the decisions made by the judge. For details see the appendix. For the state of São Paulo we use the website of the court, which contains the same information as the *Diários de Justiça* in a more structured format. Figure 1.1 shows the coverage in time of the court data by state.

Figure 1.1: Coverage of the judicial data over time by state. The different coverage by state is due to how far back the *Diários de Justiça* are available at the court websites.



The final decision of the judge comes in three main categories. Either the judge rules the case totally in favor of the prosecution (*procedente*), partially in favor of the prosecution

(*parcialmente procedente*) or acquits the defendant of all charges (*improcedente*). In our main specification we consider the latter case as the defendant winning, and a ruling partially or fully in favor of the prosecution as the defendant losing.

We estimate the date of filing of the court cases as follows: First, the filing of a case is recorded in the *Diários de Justiça*, thus if the filing occurred within the sample period, we use this as the filing date. Otherwise we know which year the case is filed from the case number, and set the filing date to June 1st the year it was filed. In order to be sure we are not miscoding cases filed after the election as cases filed before the election we exclude all cases filed in an election year for which we don't know the exact filing date.

Each time there is a publication about a case in the *Diário de Justiça* the names of the lawyers registered on the case are listed, in most states together with their unique registration number with the *Ordem dos Advogados do Brasil* (OAB). We use this information to create a data set with all the lawyers registered on each *Ação de Improbidade* at each date there is a publication about that case. From this data set we calculate *lawyer experience* by the number of previous *Ações de Improbidade* that the lawyer has worked, and *lawyer success rate* by the share of these cases in which the client of the lawyer has been acquitted, given that the case has arrived at a final decision.

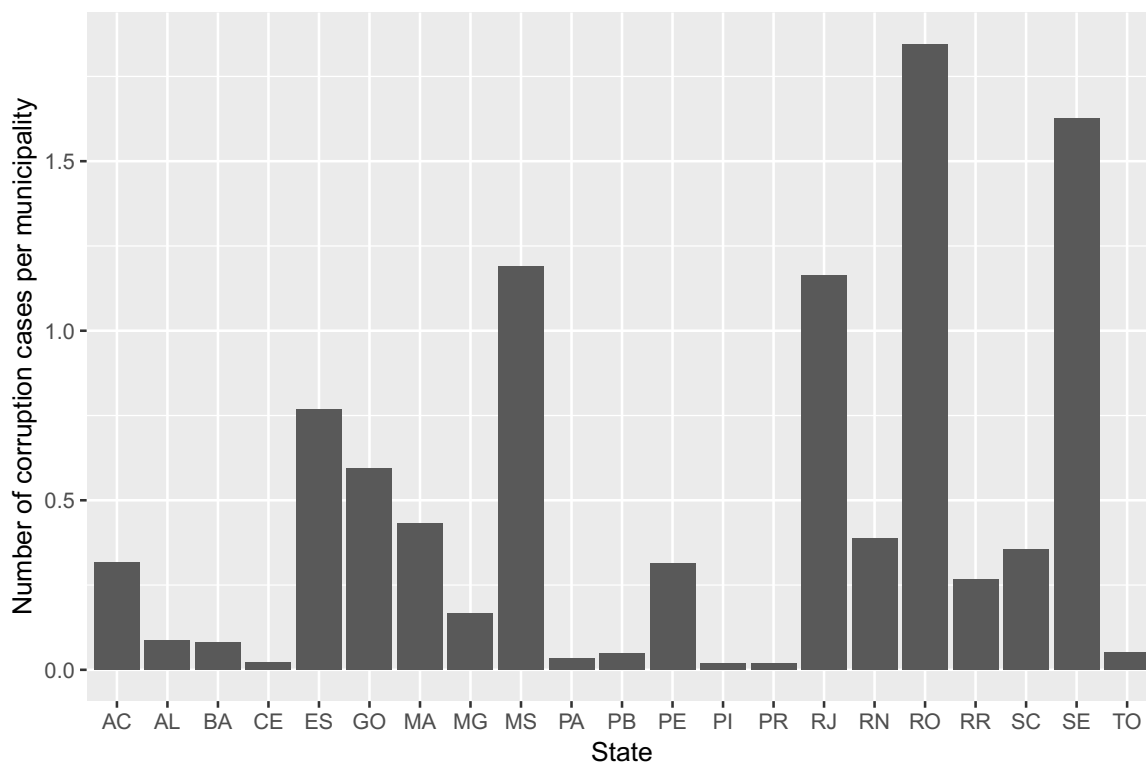
In addition, we supplement the *Diários de Justiça* with the *Cadastro Nacional de Condenações Cíveis por Ato de Improbidade Administrativa e Inelegibilidade* which is a public database of politicians who has either been convicted in an *Ação de Improbidade* and this has been upheld by an appeals court, or the possibilities for appeals have been exhausted.

1.3.2 Electoral data and matching

We use election results and candidate characteristics from the electoral authorities (Tribunal Superior Eleitoral). The candidate characteristics include information about the age, gender, education, occupation, municipality of birth, and total campaign expenditure of the candidates. We match court cases to politicians on perfect name matching. If the names of defendants include the name of a politician, ignoring accents, we code this as a match. We

exclude common names via the following procedure: Each token in a name is assigned a log likelihood based on the rate with which this token appears in the names of all litigants involved in a court case in the state of São Paulo between 2012 and 2017. A match is kept if the sum of the log likelihood of all tokens in the name is less than -30. Examples of names which are just uncommon enough to be included as matches are "Jefferson Carvalho Sales", "Eliana Aparecida dos Santos", and "Terezinha de Jesus Costa". Figure 1.2 shows the number of cases identified per municipality by state. On average we identify about one Ação de Improbidade case involving a local politician for every two municipalities.

Figure 1.2: *The number of identified Ações de Improbidade involving candidates for mayor or city council in the 2012 and 2016 elections per municipality by state.*



1.3.3 Summary statistics

Summary statistics for all identified Ações de Improbidade involving candidates in the 2012 and 2016 local elections, filed before and decided after the election, are presented in Table

Table 1.1: *Summary statistics*

Statistic	Mean	St. Dev.	N
Politician acquitted on all charges	0.471	0.500	448
Politician loses on some charges	0.228	0.420	448
Politician loses on all charges	0.344	0.475	448
Year filed	2,010.587	2.927	448
Year decided	2,015.413	1.819	448
Years between filing and decision	4.833	2.595	448
Decided during the 2012 electoral term	0.679	0.468	448
Decided during the 2016 electoral term	0.321	0.468	448
Decided before the next election	0.917	0.276	448
Decided within first two years of term	0.629	0.483	448
Court located in municipality	0.799	0.401	373
Candidate for mayor	0.576	0.495	448
Candidate for city council	0.424	0.495	448
Mayor running for re-election	0.461	0.499	258
Ex mayor	0.864	0.343	258
Elected mayor	0.461	0.499	258
Elected city councilor	0.326	0.470	190
Number of lawyers	2.618	1.986	298
Average lawyer experience	4.563	8.701	298
Average lawyer success rate	0.334	0.341	114

Note: Civil corruption cases involving candidates for mayor and city council in the 2012 and 2016 elections filed before and decided after the election. The variables ‘elected mayor’, ‘ex mayor’, ‘elected city councilor’, and ‘mayor running for re-election’ are calculated only among candidates running for the respective offices. Lawyer experience is the number of other Ações de Improbidade that the lawyer has worked on prior to the election date. Lawyer success rate is the share of these cases in which the client of the lawyer has been acquitted, conditional on there having been a final decision before the election.

1.1. Most of the cases (58%) involve candidates for mayor. Among these about 86% of the candidates are ex mayors, and 46% are incumbent mayors running for re-election. From reading a sample of the cases, most seem to involve either fraud in municipal contracting or the hiring of municipal workers without following proper procedure. The cases typically take a long time to reach a decision, on average 5 years. But 92% of the cases are decided before the next election, and 63% within the first two years after the election. The politician wins on all charges in 47% of the cases, and wins on some charges in 23% of the cases. Only in 34% of the cases does the politician lose on all charges. The court is located in the municipality for 80% of the cases. There are on average 2.6 lawyers registered on each case. On average each lawyer has experience from 4.6 previous Ações de Improbidade, but there is a very large variation in the level of experience, with the most experienced lawyer in our sample having worked with 150 previous cases.

1.4 Empirical strategy

We want to estimate the effect of political power on judicial decisions. If politicians in power are shown to be more likely to win in court than opposition politicians this does not prove that decisions are affected by the political power of the litigant, since elected politicians and politicians out of office are likely to be involved in very different types of cases. The ideal experiment would be to randomly allocate elected offices to politicians and look at the effect on judicial decisions on *already filed* cases. We were not able to run this experiment, so we are doing the second best which is to exploit close elections as a natural experiment. In particular, we look at corruption cases which are filed before and decided after the election, comparing politicians who marginally won the election with politicians who marginally lost the election. The idea is that the winner of a very close election is as good as randomly determined. Thus, winning and losing politicians should on average be involved in similar cases before the election, and any difference in judicial decisions has to be due to the outcome of the election. Formally, we use the bias-corrected estimator proposed by Calonico *et al.* (2014) with local linear regression for the estimate and local

quadratic regression for the bias-correction as our main specification. The specification for the local linear regression is

$$y_{ic} = \alpha + \beta E_i + \gamma WM_i + \delta E_i WM_i + \varepsilon_{ic}$$

where i is a politician and c is a corruption case filed before and decided after the election. The variable WM_i is the electoral win margin of the politician, and E_i is a dummy for whether the politician got elected. The outcome y_{ic} will vary, but in the baseline model it will be a dummy for whether the politician was acquitted in the corruption case. To avoid researcher discretion in the choice of control variables, we tie our hands by using no control variables. As a placebo check we also run the above regression for cases decided before the election. If close elections are indeed randomly decided, we should not see any effect of the election on these cases. We will also report the main results using conventional local linear specifications for different bandwidths, including the Imbens and Kalyanaraman (2012) optimal bandwidth. We cluster standard errors at the judicial district level.

For candidates for city council we calculate their win margin as follows. Let V_i denote the votes received by the candidate, V the total number of votes cast, and n the number of seats in the council. In the case the candidate did not receive a seat the win margin is calculated as

$$WM_i = \frac{V_i - \bar{V}_i}{V/n}$$

where $\bar{V}_i$ is the votes received by the candidate with the fewest votes within the members of i 's electoral coalition who won a seat. For candidates who won a seat the win margin is calculated as

$$WM_i = \frac{V_i - \underline{V}_i}{V/n}$$

where $\underline{V}_i$ is the votes received by the candidate with the most votes within the members of i 's electoral coalition who did not win a seat. We divide by V/n since this is the number of

Table 1.2: *Balance on pre-election variables*

Variable	Difference	p-value	Mean	Observations
Civil corruption case pending	−0.001	0.37	0.003	171623
Female	0.040	0.70	0.107	448
Incumbent mayor	−0.079	0.57	0.266	448
Farmer	0.045	0.60	0.031	448
Teacher	−0.020	0.59	0.042	448
Lawyer	−0.079	0.42	0.060	448
Civil servant	−0.049	0.66	0.112	448
Vendor	0.014	0.71	0.058	448
Businessman	−0.057	0.63	0.083	448
Higher education	0.288	0.18	0.549	448
Married	0.005	0.96	0.703	448
PMDB	0.119	0.39	0.228	448
PSDB	−0.070	0.17	0.051	448
PT	0.039	0.55	0.058	448
Number of lawyers	0.719	0.29	2.308	118
Average lawyer experience	−2.677	0.56	8.640	118
Average lawyer success rate	0.072	0.88	0.299	73

Notes: Regression discontinuity coefficients showing the estimated difference between marginal winning and marginal losing candidates in various pre-election covariates. The first row includes all candidates for mayor and city council in the 2012 and 2016 Brazilian elections. The remaining rows include only politicians in our estimation sample: Those involved in an Ação de Improbidade filed before and decided after the election. Number of lawyers, lawyer experience, and lawyer success rate calculated before the election. Estimated using the bias-corrected estimator proposed by Calonico, Cattaneo, and Titiunik (2014) with a local linear regression for the estimate and local quadratic regression for the bias-correction. The running variable is the share of votes obtained in the election. No control variables. Standard errors clustered at the judicial district level.

votes behind each seat in the council.

1.5 Main results

Table 1.2 shows results from running the main regression discontinuity specification with a wide range of pre-election covariates as outcome variables. If close elections are indeed randomly determined we should not see any systematic differences between marginal winners and marginal losers on these variables. The first row looks at all candidates in the 2012 and 2016 local elections and show that marginal winners do not have a significantly

Table 1.3: *Main regression discontinuity results*

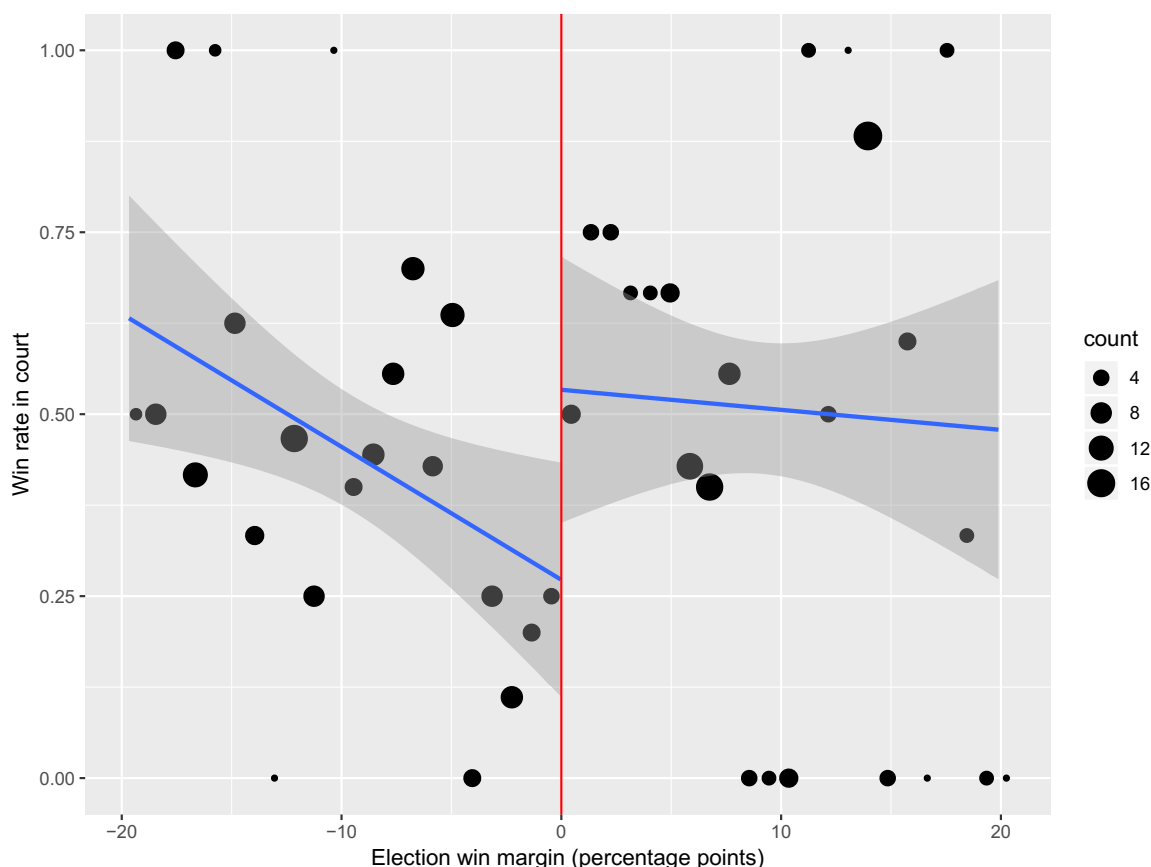
	Wins in court	Wins in court (2nd definition)	Placebo	Years since election
Elected (se)	0.648*** (0.17)	0.579*** (0.19)	0.042 (0.11)	-0.329 (0.46)
N	182	182	292	182
Bandwidth	0.12	0.13	0.18	0.13
Mean Marg. Loser	0.11	0.43	0.43	1.72

Notes: Regression discontinuity estimates using the bias-corrected estimator proposed by Calonico, Cattaneo, and Titiunik (2014) with a local linear regression for the estimate and local quadratic regression for the bias-correction. The running variable is the share of votes obtained in the election. No control variables. Civil corruption cases involving candidates for mayor and city council in the 2012 and 2016 elections filed before and decided after the election. Except for Column 3 which uses cases decided before the election. The variable 'Wins in court' is a dummy for the politician being acquitted on all charges. The variable 'Wins in court (2nd definition)' is a dummy for the politician being acquitted on all or some charges. 'Placebo' uses the same outcome as Column 1, but looks at cases decided before the election. 'Years since election' is the number of years between the election date and the final decision in the Ação de Improbidade. Mean Marg. Loser shows the estimated mean of the outcome variable for the marginal loser, using the local linear fit. Standard errors clustered at the judicial district level. $*p \leq 0.1$; $**p \leq 0.05$; $***p \leq 0.01$.

lower probability than marginal losers of being involved in an Ação de Improbidade at the time of the election. Thus, there is no evidence that that somehow the less corrupt are able to systematically win in close races. The remaining rows look at the balance within the estimation sample: Only candidates who are involved in an Ação de Improbidade at the time of the election. For none of the variables is there a statistically significant difference between marginal winners and marginal losers, consistent with the winner in close races being randomly determined.

The result of our main regression discontinuity specification is presented in Column 1 of Table 1.3. The point estimate indicates that marginal electoral winners are 65 percentage points more likely than marginal electoral losers to be acquitted from the corruption charges. The effect is statistically significant at the one percent level. Figure 1.3 shows a clear jump in the probability of being acquitted when the number of votes passes the threshold necessary to win the election. Results using the Imbens and Kalyanaraman (2012) optimal bandwidth

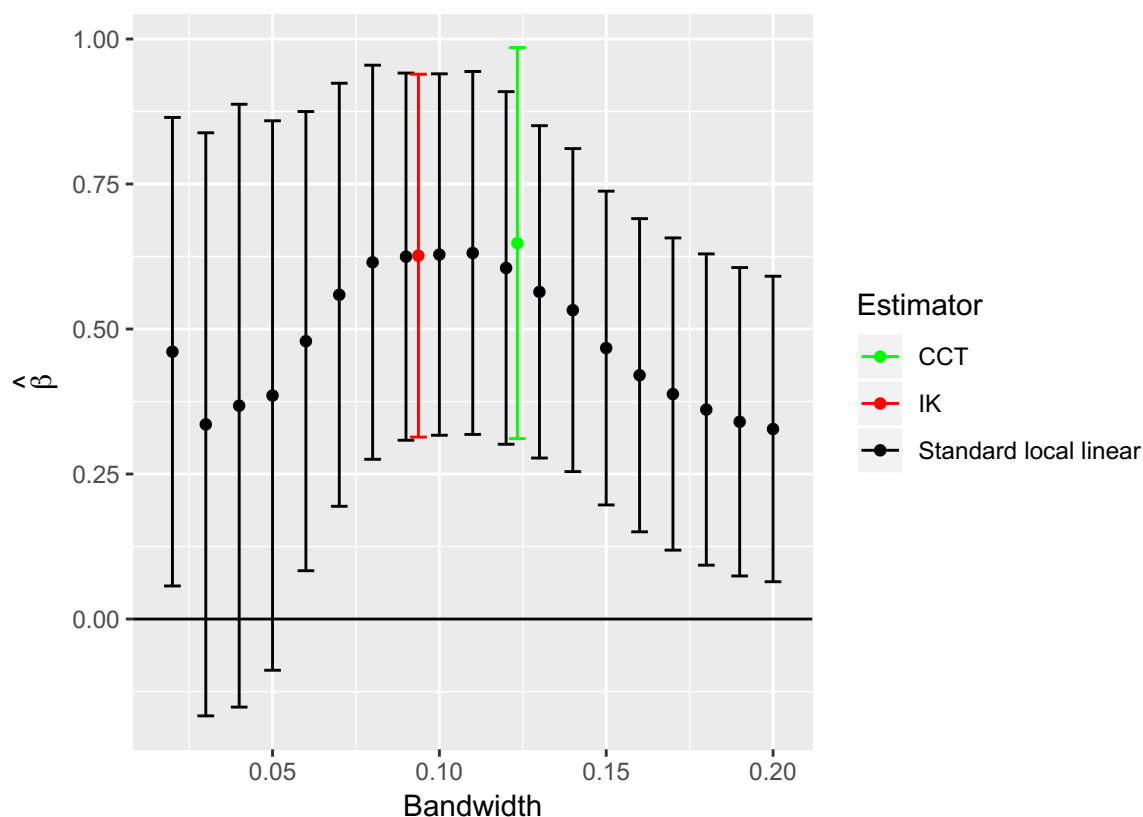
Figure 1.3: Regression discontinuity plot. *Ações de Improbidade* involving candidates for mayor or city council in the 2012 and 2016 elections filed before and decided after the election.



as well as local linear regression discontinuity specifications for different bandwidth sizes are presented in Figure 1.4. The optimal bandwidth selectors choose bandwidths of 13 and 9 percentage points, but the effect is statistically significant at the 5% level for any bandwidth larger than 5 percentage points. Column 2 of Table 1.3 shows that the result is robust to counting a ruling partially in favor of the prosecution as an acquittal rather than a conviction. The result of the placebo test is presented in Column 3. Reassuringly, there is no effect of winning the election on cases decided *before* the election. The absence of any jump at the threshold for cases decided before the election can be visually inspected in Figure 1.5.

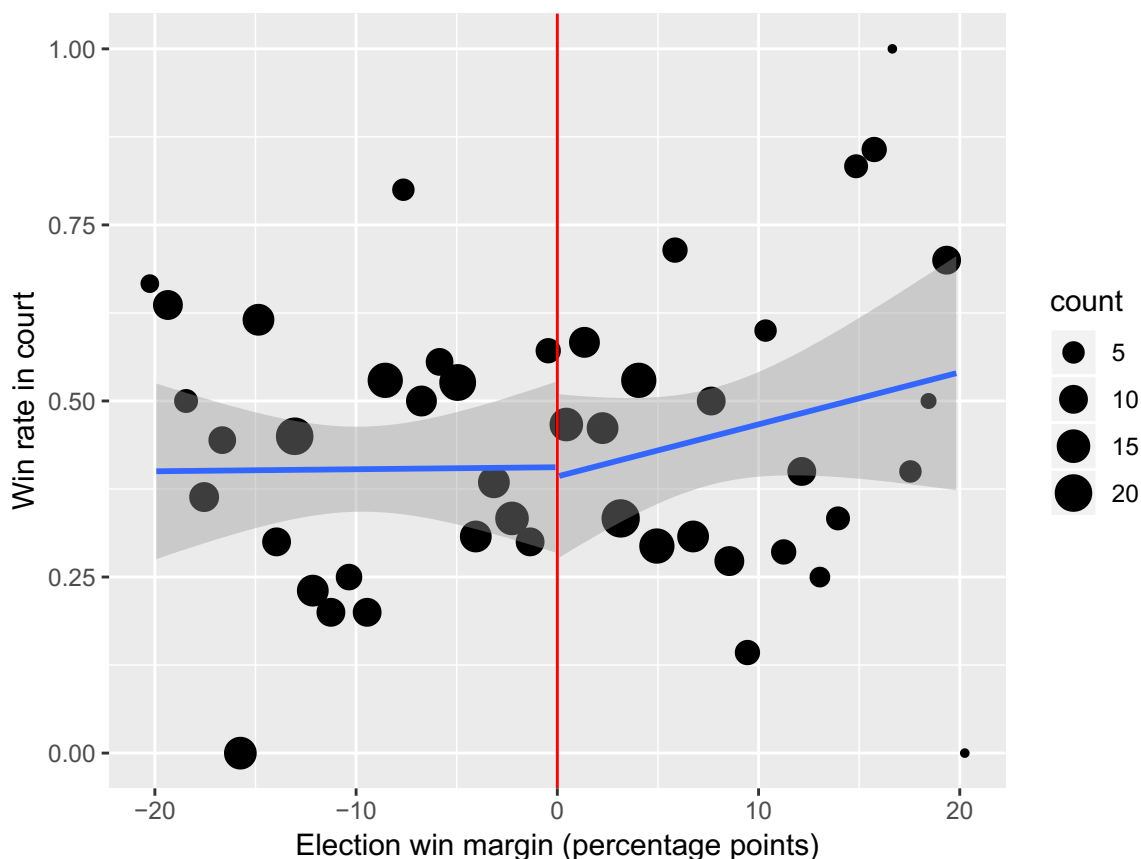
One possible explanation of the result is that it might be driven by judges postponing unfavorable decisions against elected politicians, not to any change in actual decisions.

Figure 1.4: Regression discontinuity estimates for different bandwidths. *Ações de Improbidade* involving candidates for mayor or city council in the 2012 and 2016 elections filed before and decided after the election. The local linear specifications uses a triangular kernel. 'IK' uses the Imbens and Kalyanaraman (2012) optimal bandwidth. 'CCT' uses the bias-corrected estimator proposed by Calonico et al. (2014) with a local linear regression for the estimate and local quadratic regression for the bias-correction. 95% confidence intervals. Standard errors clustered at the judicial district level.



While this would also be evidence of elected politician receiving a favorable treatment by the judiciary, it would lead to a different interpretation. In order to test for whether elected politicians are able to postpone decisions in their cases, we run regression discontinuity regressions with the years between filing and the decision in Column 4 as outcome variable. The point estimate indicates that cases involving marginal winners are in fact decided four months *faster* than cases involving marginal losers. However, the effect is statistically insignificant. Furthermore, the McCrary plot in Figure 1.6 does not reveal any clear decrease in the number of cases as the vote share crosses the threshold for being elected. Thus, the

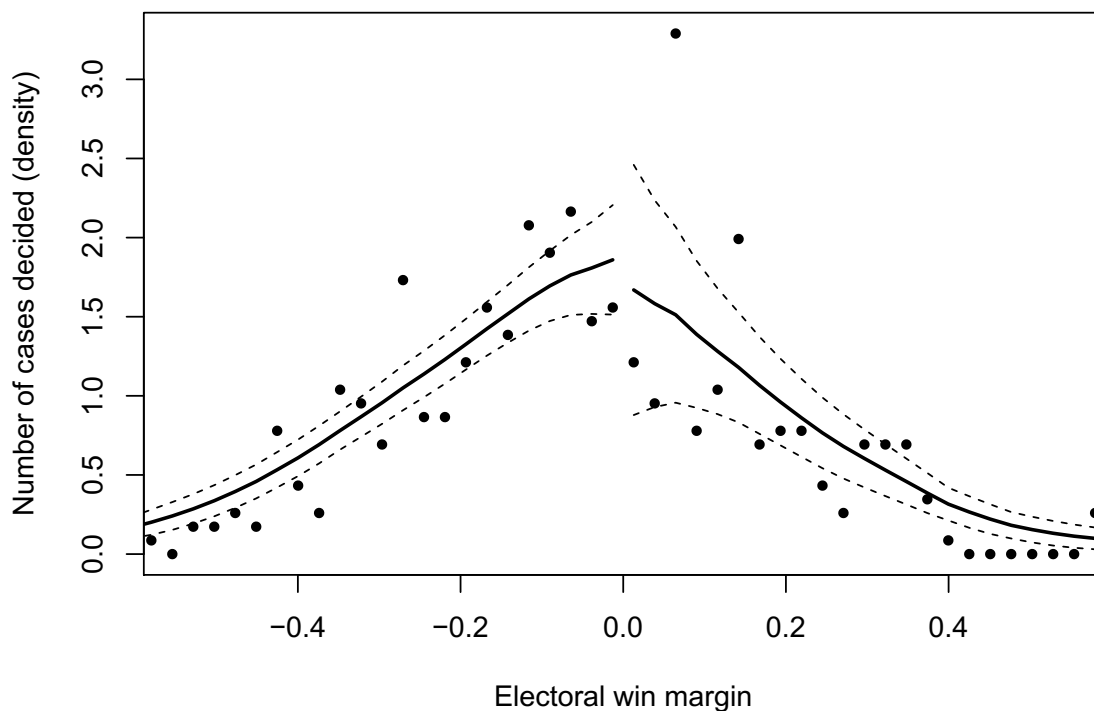
Figure 1.5: *Placebo regression discontinuity plot. Ações de Improbidade involving candidates for mayor or city council in the 2012 and 2016 elections decided before the election.*



result is unlikely to be explained by judges postponing decisions against elected politicians.

In Panel A of Table 1.5 we present some basic heterogeneous effects, where we split the sample into subsamples and estimate the main regression discontinuity specification separately on each subsample. We estimate p -values of the difference in effects between subsamples under the assumption that the subsamples are independently drawn. The first two rows shows that the effect is driven by both candidates for mayor and city council, with the effect being larger for mayoral candidates. The next rows show that there is no statistically significant difference in the effects comparing large and small municipalities and comparing incumbents with non-incumbents, although the point estimate is larger for large municipalities.

Figure 1.6: McCrary sorting test. *Ações de Improbidade* involving candidates for mayor or city council in the 2012 and 2016 elections filed before and decided after the election.



1.6 Mechanisms

Why are Brazilian local politicians more likely to be acquitted of corruption charges if they win the election? In this paper we focus on two distinct types of mechanisms. First, it could be that the electoral winners are able to be acquitted at a higher rate by hiring lawyers that are better at convincing the judge of their innocence. This could be due to knowing the law better, being more skillful in the use of interlocutory appeals, or just from being a great orator. There are at least three reasons to believe that electoral winners might have better lawyers: They might have more to lose if they are convicted, as they might have to step down from their office; they might be less liquidity constrained due to their official salary or other new sources of income; and they might be able to use the lawyers of the municipality to

receive legal advice. The latter is illegal in Brazil since Ações de Improbidade are considered as private lawsuits of the politician, however we know anecdotally that this might happen.

The second type of mechanism we will consider is that agents of justice such as judges, judicial staff, prosecutors, and witnesses, might be influenced by the power of the politician through channels unrelated to the legal process, which we call judicial subversion. It might be through the politician's control over public resources. For instance, he can make it difficult for someone who witnesses against him to obtain a job in the municipality or to receive municipal contracts. Judges and prosecutors themselves are prohibited from receiving government jobs or contracts, but there are no rules preventing a mayor from offering positions in the municipality to their relatives or friends, maybe with the implicit message that they will continue in the jobs as long as the law enforcers acts friendly towards the mayor. Also, both the prosecutor and the judge is required to live in the judicial district (though some are granted an exemption) and rely on municipal services such as water, electricity, and health care. Furthermore, there might be social benefits from siding with an elected politician such as being invited to dinner parties with the city elite. Finally, it could be that law enforcers are purely psychologically influenced by the power of the politician without expecting any rewards or punishments.

In this paper we are not going to be able to say much about which of these non-legal ways that the power of the politician might influence the outcome of the court case. But what we believe we are able to is to disentangle this broad category of mechanisms, from the ways in which power might affect judicial decision through the legal process by the work of more skillful lawyers. Although knowing which of the non-legal ways of influence are at play would be very interesting, we think that being able to disentangle between these two main categories of mechanisms is also valuable. The reason is that the policy response might be very different depending on which of the two types of mechanisms are driving the result. If the effect is through a direct influence of political power over law enforcers a reasonable policy would be to move the trial to a different location outside the reach of the politicians power, maybe to a neighboring judicial district. Conversely, if the effect is

due to elected politicians hiring better lawyers such a policy will not work. Instead a more effective policy might for instance be to provide legal assistance to opposition politicians not able to afford a high quality lawyer.

In the following subsections we will show several tests which indicate that the first type of mechanism, elected politicians having better lawyers, is unlikely to explain the result, and that the evidence is more consistent with the effect coming from the second type of mechanism, elected politicians influencing judicial outcomes via non-legal means. In Section 1.6.1 we will show that there is no evidence that marginal electoral winners change to having more experienced or successful lawyers registered on their case after the election at a higher rate than marginal losers. Section 1.6.2 shows that the effect is larger if the court is located in the municipality than if the court is in a neighboring municipality. In Section 1.6.3 we show that the effect seem to be larger if the decision is made early in the electoral term. In Section 1.6.4 we argue that there is no reasons to believe that marginal winners have more at stake in the case, and thus be willing to hire better lawyers than marginal losers. Finally, in Section 1.6.5 we show that there is no evidence that the effect is larger for more liquidity constrained politicians.

1.6.1 Direct tests of the lawyer hypothesis

The most direct test of whether the result is driven by elected politician having better lawyers is to measure whether marginal electoral winners tend to register more or better lawyers on their case after the election, at a higher rate than marginal losers. This is possible to check since at any date that there is a publication about a given case in the *Diários de Justiça* we know the identity of the lawyers on the case. On average each case in our sample has 5 publications, with 70% of the publications being after the election. Thus, counting the *number* of lawyers and whether there are new lawyers added to the case at each date is straight forward. As proxies for the quality of each lawyer we use *lawyer experience* defined as the number of other *Ações de Improbidade* that the lawyer has worked on prior to the election, and *lawyer success rate* defined as the share of these cases in which the client of

Table 1.4: *The effect of winning the election on the quantity and quality of lawyers*

	Difference between after and before election in:			
	Number of lawyers	Number of new lawyers	Average lawyer experience	Average lawyer past success
Elected (se)	0.0057 (0.441)	-0.0657 (0.326)	0.2096 (1.507)	-0.0452 (0.054)
N	86	86	86	42
Bandwidth	0.13	0.24	0.14	0.11
Mean Marg. Loser	0.34	0.12	-0.89	-0.0014

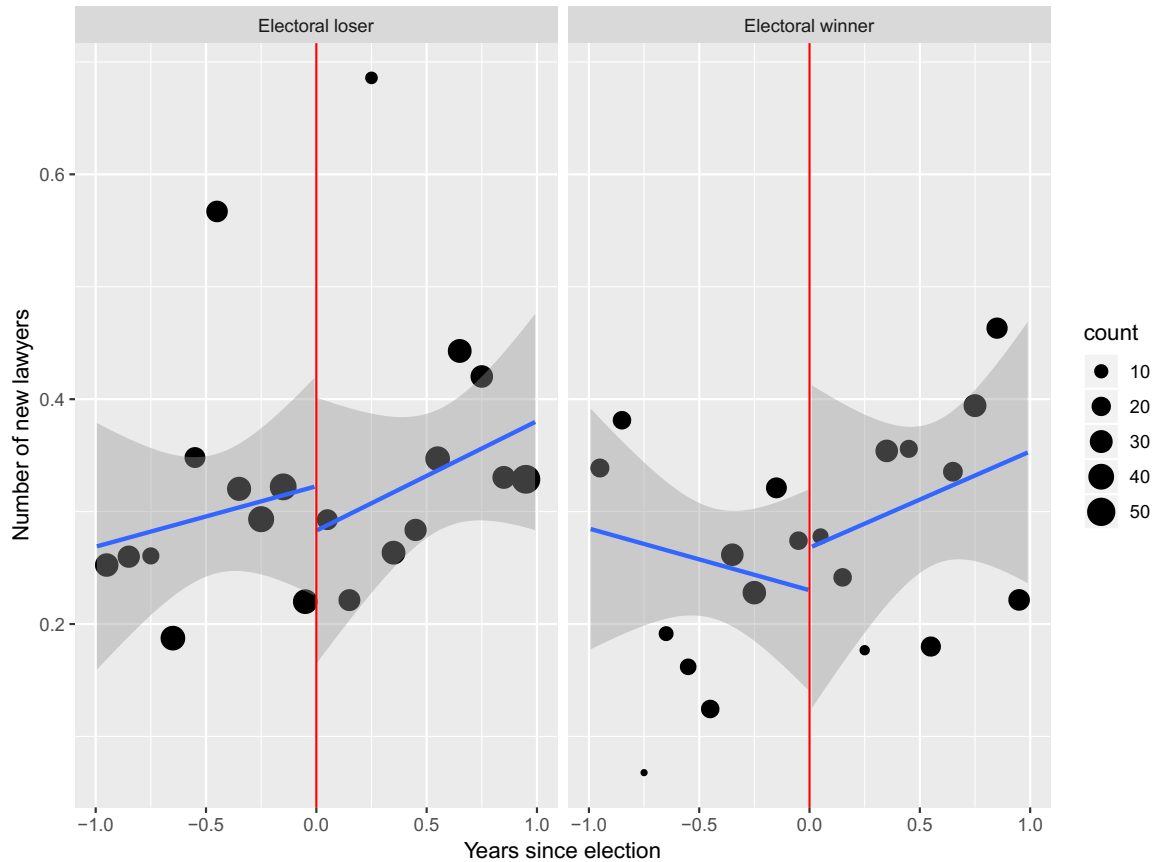
Notes: Regression discontinuity estimates where the outcome variable is the difference between the average of the respective variables across all publications made in the Diário de Justiça after the election and the same average before the election. Only cases where we have publications in the Diário de Justiça both before and after the election. The experience of a lawyer is defined as the number of other civil corruption cases she has worked on prior to the election. Her past success is the rate at which her client was acquitted in civil corruption cases decided before the election. Coefficients estimated using the bias-corrected estimator proposed by Calonico, Cattaneo, and Titiunik (2014) with a local linear regression for the estimate and local quadratic regression for the bias-correction. The running variable is the share of votes obtained in the election. No control variables. Civil corruption cases involving candidates for mayor and city council in the 2012 and 2016 elections filed before and decided after the election. Standard errors clustered at the judicial district level.

* $p \leq 0.1$; ** $p \leq 0.05$; *** $p \leq 0.01$.

the lawyer has been acquitted, conditional on there having been a final decision in the case prior to the election. We calculate the average number and quality of lawyers in Diário de Justiça publications about the case before and after the election, respectively. To test whether marginal electoral winners tend to hire better lawyers after the election compared to marginal losers we run the main regression discontinuity specification with the outcome being the increase in lawyer quantity or quality after the election. If the result is explained by marginal electoral winners having better lawyers, we would expect there to be a clear jump in the post-election increase in lawyer quality or quantity for politicians who obtain just enough votes to cross the threshold for being elected.

The results are presented in Table 1.4. The point estimate in Column 1 shows that marginal electoral winners increase the number of lawyers by only 0.6 percentage points more than marginal losers. The standard error is large, so we cannot use this result to claim that electoral winners in general do not have more lawyers, but at least in our sample there

Figure 1.7: The number of new lawyers over the electoral cycle. One observation per publication in the *Diário de Justiça*. *Ações de Improbidade* involving candidates for mayor and city council in the 2012 and 2016 elections filed before and decided after the election.



does not seem to be any clear difference between the post-election increase in the number of lawyers of marginal winners and marginal losers. Similarly, Column 2 shows it is not the case in our sample that marginal winners are hiring new lawyers after the election at a higher rate than marginal losers. Also, Figure 1.7, which plots the number of new lawyers over time, does not reveal any tendency for neither electoral losers nor electoral winners to hire new lawyers at a higher rate after the election than before. There are also very small differences in average lawyer experience and success rates. The point estimate in Column 3 of Table 1.4 indicates that the pool of lawyers working for the marginal winner decrease by on average having experience from 0.9 *Ações de Improbidade* less, with the decrease for marginal winners being 0.7. This is a small difference compared to the average lawyer

experience of 4. Finally, Column 4 shows that the post-election change in the rates at which the lawyers have won the case for their clients in previous Ações de Improbidade are very similar for marginal winners and marginal losers. In sum, it does not seem to be the case that marginal winners increase the quantity or quality of the lawyers who *formally* work on their case after the election any more than marginal losers.

While it is tempting to rule out the lawyer hypothesis completely after this evidence, there might still be changes to the quality of legal counsel that just looking at the lawyers formally registered on the cases does not detect. First, it might be the case that electoral winners receive informal help from lawyers not registered on the case. This is particularly an issue if elected politicians receive help from municipal lawyers, since these are not allowed to work on the case. Second, it might be the case that electoral winners pay existing lawyers to work longer hours, without changing the identity of any of the lawyers. Thus, in the following subsections we will also show a series of other, less direct, tests.

1.6.2 The location of the court

Brazilian judicial districts are often composed of several municipalities with the court typically seated in the largest municipality. This means that some of the politicians in our sample are tried in a court that is located in their municipality, while others are tried in a court located in a neighboring municipality. Normally law enforcers, including the judge and the prosecutor, will live in the municipality where the court is located, both because this gives a shorter commute and might have more amenities since it typically the largest of the municipalities composing the judicial district. Consequently, the chances of a politician of influencing judicial outcomes by offering easier access to public services or municipal jobs and contracts to law enforcers, their friends and families, will likely be much larger if the court is located in the municipality where the politician holds power. In other words, if the marginal winners are favored by law enforcers due to influencing decisions via non-legal means we would expect our main regression discontinuity estimate to be larger if the court is located in the municipality where the politicians are running for office. Conversely, if the

Table 1.5: *Heterogeneous effects*

		Coef.	(se)	N	Band- width	p-value of difference
A: General						
Candidate for Mayor	Yes	1.15***	(0.29)	119	0.10	0.09
	No	0.48*	(0.27)	63	0.14	
Municipality population larger than median	Yes	0.90***	(0.11)	83	0.08	0.25
	No	0.53*	(0.30)	99	0.12	
Incumbent	Yes	0.71**	(0.28)	101	0.10	0.66
	No	0.53*	(0.30)	99	0.12	
B: Main tests						
The court is located in the municipality	Yes	0.78***	(0.22)	130	0.10	0.06
	No	−0.27	(0.52)	25	0.10	
Court case decided in the first two years of term	Yes	0.86***	(0.18)	76	0.11	0.20
	No	0.40	(0.32)	82	0.14	
C: Liquidity constraints						
Politician has higher education	Yes	0.64***	(0.17)	110	0.10	1.00
	No	0.64**	(0.31)	72	0.13	
Politician has a larger than median campaign	Yes	1.14***	(0.27)	78	0.09	0.20
	No	0.70***	(0.22)	98	0.11	

Notes: Regression discontinuity estimates for different subsamples. '*p*-value of difference' is the *p*-value of the difference in estimated effects between the two subsamples, assuming that the two subsamples are independently drawn. 'Incumbent' includes both incumbent mayors and city councillors. 'Court case decided in the first two years of term' is whether the final decision in the Ação de Improbidade was made within two years after the date of the election. Coefficients estimated using the bias-corrected estimator proposed by Calonico, Cattaneo, and Titiunik (2014) with a local linear regression for the estimate and local quadratic regression for the bias-correction. The running variable is the share of votes obtained in the election. No control variables. Civil corruption cases involving candidates for mayor and city council in the 2012 and 2016 elections filed before and decided after the election. Standard errors clustered at the judicial district level. * $p \leq 0.1$; ** $p \leq 0.05$; *** $p \leq 0.01$.

effect is driven by elected politicians having access to better lawyers, there are no reason to believe that there should be a differential effect. Whether a good lawyer is able to convince the judge, should not depend on the location of the court.

The result of this exercise is presented in the first row of Panel B in Table 1.5. For cases which are tried in the municipality where the politician is running for office, the effect of winning the election on the probability of winning in court is 78 percentage points and significant at the 1 percent level. On the other hand, the estimated effect is -27 percentage points and statistically insignificant for cases which are tried in a court located in a different municipality. The difference between the two effects has a p -value of 6%. This means that the result is unlikely to be due only to elected politicians having better lawyers. It is also evidence against *some* purely psychological explanations of our result, such as judges favoring politicians in power due to not feeling legitimated to rule against democratically elected politicians in general.

1.6.3 Cases decided early versus late in the electoral term

Some politicians might be involved in an Ação de Improbidade which is already at an advanced stage at the time of the election, meaning that a final decision in their case is close. Other politicians might be involved in cases which are in their early stages, and a decision is only made towards the end of the electoral term when the winner is about to leave office and new elections are upcoming. Is the effect largest for the first types of cases or the latter? This is an interesting question since the two types of mechanisms have opposing predictions. If the result is driven by electoral winners having better lawyers, we should expect the effect of winning the election on judicial outcomes to be larger if the case it at an early stage when the politician takes office, since it gives lawyers more time to have an impact on the final decision. On the other hand, we would expect a small effect for cases decided right after the politician takes office since most of the lawyer work has already been done before the election. However, if the result is driven by law enforcers hoping to be rewarded for a friendly attitude towards the politicians in power with easier access to public services or

municipal jobs and contracts, or fearing reprisals, we would expect the result to be largest if the decision is made just after the politician takes office. It is just after taking office that there are most future opportunities for the elected politician to use his power to reward friendly law enforcers or punish those that are disloyal. If, on the other hand, the decision in the case is made towards the end of the electoral term, the marginal winner might soon be out of office and the marginal loser might be the one who is most able to offer promises of future rewards and punishments, due to the incumbency disadvantage to be discussed below.

In order to test whether the effect is largest if the case is decided early or late in the electoral term, we run our main regression discontinuity specification on all cases that were decided in the first two years after the electoral winner took office, and all cases decided in the last two years of the term, respectively. The result is presented in Panel B of Table 1.5. The regression discontinuity point estimate is largest if the case is decided early in the term: A marginal electoral winner is estimated to have a 86 percentage points greater chance of being acquitted than a marginal loser if the case is decided within the first two years of the term, but only 40 percentage points greater chance of being acquitted if the case is decided in the last two years. However, the difference between the two effects has a p -value of 0.2, meaning that we can only be 80% confident that the effect is actually larger in the first two years. Another caveat of this result is that we are splitting the sample on a potentially endogenous variable, the timing of the decision. While the evidence is consistent with the effect being due to power influencing judicial outcomes by non-legal means, it could also be consistent with the lawyer hypothesis if for instance cases in which marginal electoral winners are expected to lose are postponed.

1.6.4 Do electoral winners have more at stake?

One of the main reasons to believe that electoral winners hire better lawyers is that they risk losing their office if they lose the case. With more at stake, they might be willing to pay more for legal counsel than electoral losers, which have no office to lose. In this section we

will argue that *marginal* electoral winners are unlikely to have better lawyers than *marginal* losers due to this reason, and that in fact marginal losers have as much at stake in case as the marginal winners. The main reason is that even if the judge invokes the penalty of loss of office or the loss of political rights, the politician can continue in office as long as there still are possibilities to appeal the decision. The Brazilian legal system is composed of four instances, the trial courts, the appeal courts, the superior court, and the supreme court, and allows for a myriad of ways to appeal both interlocutory and final decisions. This, combined with the slowness of the judicial system, means that it is unlikely that an electoral winner will have to step down before his term is over even if the penalty of loss of office is imposed by the trial court in the first year of office. In fact by matching our data with the *Cadastro Nacional de Condenações Cíveis por Ato de Improbidade Administrativa e Inelegibilidade* which keeps track of convictions in Ações de Improbidade for which all possibilities of appeals has been exhausted, we found that no electoral winner in our data has actually had to step down during the term due to an Ação de Improbidade.

While it is very unlikely that any elected politician has to step down in the same term as the trial court decision, receiving the penalty of loss of political rights in an Ação de Improbidade can have serious consequences for the *future* political career of a politician. This was especially true after the passing of the *Clean Record Act*) (Lei da Ficha Limpa) in 2010, which stipulates that if such a sentence has been confirmed by the appeals court the politician is barred from running for any electoral office for the next eight years, even when possibilities for appeals has not been exhausted. However, this consequence applies to electoral losers as well, since marginal electoral losers also might care about their future political career. In fact, since there is an incumbency *disadvantage* among mayoral candidates in Brazil Klašnja and Titunik (see e.g. 2017) the marginal loser has a higher chance of becoming elected in the next election, and thus might even have *more* at stake in terms of a future political career than the marginal electoral winner. In Section A.2 in the Appendix we show that the marginal loser in Brazilian mayoral races is 16 percentage points more likely to become the next mayor than the marginal winner, and 13 percentage points more

likely to be elected mayor in the election after. On average, the expected future years in any political office, including state and federal offices, is 1.2 years larger for marginal losers than for marginal winners. There are no differences in the future political careers between marginal winners and marginal losers in elections to the city council.

1.6.5 Liquidity constraints

Even if marginal winners and marginal losers might have the same *willingness* to pay for a great lawyer due to the argument above, it could be the case that the *ability* to pay is larger for the marginal winners. That is, marginal winners might be less liquidity constrained. The official salary and other income that might derive from holding electoral office could make electoral winners more able to pay for lawyers. In order to test whether liquidity constraints could explain the result we investigate in this section whether there are heterogeneous effects depending on whether politicians are likely to be liquidity constrained based on pre-election covariates. The two characteristics we look at are whether the politician has higher education, and whether the campaign expenses of the politician is larger than the median campaign. The results are presented in Panel C of Table 1.5 which shows no evidence of there being a larger effect for more liquidity constrained politicians. The point estimate is identical for politicians with and without higher education, and larger for politicians with large campaigns. Thus, elected politicians having better lawyers due to liquidity constraints is unlikely to be the main driver behind the result.

1.6.6 Which law enforcers are influenced?

The effect might be driven by judges, prosecutors, judicial staff, or witnesses being influenced by political power. Which of these actors might be driving the result? While we do not have any evidence which can disentangle whether the effect comes from the judge, judicial staff, or witnesses, it seems unlikely that the result is driven only by prosecutors being influenced due to two reasons. First, a natural way for public prosecutors to favor an elected politician would be to ask the judge to dismiss the case on a technicality. However, the results

presented in Section 1.5 did not show a large decrease in the number of cases involving electoral winners after the election, which we would expect to see if such dismissals are going on. Second, the biggest impact that a prosecutor can have on weakening a case is during the investigation, where they have a lot of discretion about which steps to take in the investigation, and whether to file a case or not. The fact that the effect seems to be larger for cases decided early in the electoral term also suggest that the effect is at least not only due to prosecutors being influenced.

1.6.7 Other mechanisms

In this section we consider two other potential mechanisms that do not fit into any of the two main categories above.

First, in the Brazilian setting mayors have special privileges when it comes to criminal court cases called *foro especial por prerrogativa de função*, colloquially known as "foro privilegiado": Instead of being tried at the trial court criminal cases involving mayors go straight to the appeals court. Many times the acts a politician is accused of in an Ação de Improbidade are criminal acts, which means that there might be a criminal case investigating some of the same facts running in parallel. If a candidate wins the election any criminal case in the trial court is sent to the appeals court, making it potentially more difficult for the prosecutor in the Ação de Improbidade. Also, the judge is required to take into consideration the decision in the related criminal case, which might be different if the case is tried in the appeals court. This might explain parts of our result. However, we believe that it cannot explain the full result. First, we have also collected criminal cases involving candidates for mayor and city council in the state of São Paulo, and there are four times as many Ações de Improbidade as there are criminal cases. Thus, even if each criminal case is related to an Ação de Improbidade this channel alone cannot explain an effect with the magnitude of 60 percentage points. Also, it cannot explain the result for city councilors who have no special privileges, and has a hard time explaining the fact that the effect is bigger if the court is located in the municipality and that the effect is larger if the case is decided earlier in the

term.

Second, it might be that elected politicians somehow are able to bribe law enforcers using their own money, rather than by using their power over the allocation of public resources. We believe that this mechanism is unlikely to explain the result, since it is not clear why marginal losers will not be able and willing to pay the same bribe as marginal winners given the discussion in Section 1.6.4 and 1.6.5. Also, cash bribes would not predict that the effect is bigger if the court is located in the municipality, since a politician should be equally able to bribe law enforcers in a neighboring municipality.

1.7 Incentives for corrupt politicians to run for election

Our main result shows that electoral winners are more likely to be acquitted from corruption charges than electoral losers. This fact could potentially lead to an adverse selection of politicians in electoral offices, as politicians who are facing corruption charges might seek electoral office to protect themselves from the law. In this section we look at whether there is any evidence consistent with such an effect. In order to test for this we construct a balanced panel of all politicians who ran for the office of mayor in the 2000, 2004, 2008, 2012 or the 2016 local elections. For each politician and election year we create a dummy variable indicating whether the public prosecution filed an Ação de Improbidade involving the politicians within the four years leading up to the election. We use this data set to investigate whether it is the case that a politician who becomes implicated in an Ação de Improbidade is more likely to run in the upcoming election, potentially in order to influence law enforcers to act more leniently. We do not have a source of random variation in whether a corruption case is filed, so we cannot provide a causal estimate of being implicated in an Ação de Improbidade on future electoral outcomes. However, we do our best to control for differences in the probability of running in future elections due to observable variables. Formally, we estimate the following regression

$$y_{it} = \alpha_i + \beta F_{it} + \gamma F_{it} * T_{it} + \mu_t + \delta_i t + \lambda_i t^2 + \eta' X_{it} + \varepsilon_{it}$$

where i refers to a politician and t to an election year. The outcome variable y_{it} will be whether the politicians runs or win in the upcoming mayoral election, and F_{it} is an indicator for whether there has been an Ação de Improbidade involving the politician filed since the last election. We add an interaction of F_{it} with the time between the election and the filing, T_{it} , to see if there is any differential effect depending on whether the case is filed closer to the election. We control for politician fixed effects, election year fixed effects, and quadratic politician specific time trends. We also include additional controls for the politician's previous political career, X_{it} . These controls include whether the politician ran, became elected, and the win margin in all previous elections back to 2000.

Table 1.6: *The filing of corruption cases and future elections*

	<i>Dependent variable:</i>	
	Running in next mayoral election	Elected mayor in next election
	(1)	(2)
Civil corruption cases filed	0.078*** (0.024)	0.019 (0.025)
Years between election and filing	−0.012 (0.010)	−0.006 (0.009)
Politician FE	Yes	Yes
Election year FE	Yes	Yes
Quadratic politician trends	Yes	Yes
Electoral controls	Yes	Yes
Observations	342,604	342,604
R ²	0.074	0.052

Notes: All candidates for mayor in the 2000, 2004, 2008, 2012, and 2016 elections. Balanced panel, including also years the politician did not run. 'Civil corruption cases filed' is a dummy for whether a civil corruption case involving the candidate was filed within four years before the election. 'Years between election and filing' is zero if no case was filed within those four years. Electoral controls are whether the candidate run, became elected, and the win margin for all of the previous elections back to 2000. Standard errors clustered at the politician level. * $p \leq 0.1$; ** $p \leq 0.05$; *** $p \leq 0.01$.

The results are presented in Table 1.6. Column 1 shows that politicians are 8 percentage points more likely to run for the office of mayor after being implicated in an Ação de

Improbidade, than what would otherwise be predicted by their earlier political career. The estimate is statistically significant at the one percent level. Column 2 shows that politicians are two percentage points more likely to become the next elected mayor after having an Ação de Improbidade filed against them, but the coefficient is not statistically significant. Since we do not have a source of exogenous variation in whether a politician is implicated in an Ação de Improbidade, these results are purely suggestive. However, the results are at least consistent with the view that a judicial system favoring elected politicians might attract corrupt politicians to electoral offices.

1.8 Implications and future research

In this paper we have estimated that Brazilian local politicians are 60 percentage points more likely to be acquitted of corruption charges if they win the election with a small margin. This is true even though local politicians have no influence over the careers paths of neither judges, prosecutors, nor judicial staff. The most likely channel is that law enforcers are influenced by the political power which comes with electoral offices, rather than from electoral winners having better lawyers. The main implication of this result is that we need more than formal separation of powers between the judiciary and the executive to ensure that the law is blind to the political power of defendants. Even with all the standard rules protecting the independence of the judiciary in place, we find that Brazilian mayors and city councilors receive a lighter justice than opposition politicians. We cannot argue that the formal checks has no bite, but at least they are not enough to create a system of justice that is totally immune to political power. Is there anything that can be done to limit political influence further? Probably. However, this study is not designed to shed light on this question, and which institutional designs that can make sure elected politicians and the opposition receive a more equal treatment by the law beyond the standard rules must be left for future research. There is, however, one policy that our results suggest that should be investigated further: To try politicians in elected offices in courts located outside of the reach of their power.

In addition to investigating institutional designs to improve further upon judicial independence, we see several directions in which to take this line of research further. First, while the evidence we present are highly suggestive we have not showed any direct evidence of law officials being benefited or harmed by the politician. A natural next step, while difficult, is to try to measure these benefits or punishments. For instance, is it the case that judges or prosecutors tend to have relatives employed in the municipality and that this happen at a higher rate if a politician accused of corruption marginally wins the election? A second line of inquiry is to exploit the random assignment of cases to judges to try to get at whether certain judges tend to favor elected politicians at a higher rate than others.

Chapter 2

Judges, Appeals and Judicial Decision Making¹

2.1 Introduction

A key feature of existing judicial systems is the possibility to appeal trial court decisions to an appellate court for reassessment. The commonly stated rationale for an appeal system is to correct errors and deviant judicial decisions in trial courts (Shavell, 1995). However, the appeal process can also be abused by opportunistic litigants, who may seek to postpone punishment or just find it a worthwhile gamble to get their case tried again. This could lead to unnecessary delays in the provision of justice and burden public resources.² Understanding which forces are causing appeals is important, especially since different judicial policy responses could be required to counter different drivers of appeals. For instance, if appeals are primarily driven by opportunistic litigant behavior, then a natural policy response is to make appeals more costly or impose limitations on appealing. On the contrary, if appeals are primarily driven by judicial errors or deviant judge behavior, then

¹Co-authored with Manudeep Bhuller

²See for instance Martineau 1984 who discusses the contribution of frivolous appeals and abusive appeal tactics to the workload of federal courts of appeals in the U.S.

policy responses should rather be aimed at limiting such behavior.

An ideal experiment to test whether appeals are *only* driven by opportunistic litigant behavior as opposed to judicial errors or deviant judge behavior is to assign identical cases to different judges and assess whether different judges are later appealed at different rates. If the share of appeals differ across judges, one could conclude that appeals must be, at least partly, driven by judicial errors or deviant judicial decisions. While we almost never see different judges deciding identical cases, we can get close to this ideal experiment by exploiting settings where court cases are randomly assigned to judges and test whether judges *on average* tend to see their decisions appealed at different rates. Recent scholarship has been exploiting such random assignment of cases to document systematic differences across trial court judges along many domains of judicial decision making.³ In this paper, we apply this research design to a novel data set that allows us to link trial court decisions to outcomes in appellate courts to investigate whether there are systematic differences in appeal tendencies across judges. The random assignment ensures that the characteristics of cases or defendants are uncorrelated to characteristics of judges, so that systematic differences in the rates at which decisions are appealed across judges can be attributed to differences in their decision making, and not to case or defendant characteristics.

We uncover striking differences in appeal rates across trial court judges. While the average appeal rate is 15%, one out of ten judges have their decisions appealed at a rate of 21% or more. Importantly, we show that the leave-out judge appeal tendency for the assigned judge is highly predictive of appeals and decision reversals in appellate courts. Our findings indicate that 20% of the variance in observed appeal rates across judges can be attributed to systematic judge differences in appeal tendencies. While judge appeal tendency is positively correlated with the fraction of cases a judge decides with an incarceration sentence, we show that systematic differences in appeal rates between judges remain after

³For instance, Bhuller *et al.* (forthcoming) document systematic variation in incarceration rates between judges, and Abrams *et al.* (2012) document systematic variation in the treatment of race between judges. To best of our knowledge, there is no existing evidence on systematic differences in appeal tendencies across trial court judges, and none of these studies have focused on the appeal process.

controlling for the rate at which the judge convicts, incarcerates, and her average sentence length. Furthermore, we estimate that 39% of the variance in appeal rates are due to other factors than these judge stringency measures.

In the second part of the paper we investigate determinants of the large differences in the rate at which different judges have their decisions appealed and reversed. We consider three reasons for why certain judges might have higher appeal tendencies, beyond their stringency: (1) they spend less time on each case and thus commit more errors; (2) they have judicial preferences which deviate from their peers; and (3) they are less experienced. We measure the speed at which judges make decisions by the average time between filing and decision for cases decided by the judge, and experience by the time since their first observed decision in our data. In order to get a measure of judicial preferences we estimate a discrete choice model for each judge. We feed the model with a large set of case and defendant characteristics and use machine learning to select the variables which are most predicting incarceration for each judge. To get a measure of how deviant the preference parameters of a judge is, we calculate the probability that these preferences will lead the judge to a different conclusion in a randomly drawn case than would a randomly drawn judicial peer.

The results show that a case is more likely to be appealed if assigned to judges with deviating judicial preferences, to fast judges, and to judges who has stayed a long time in the same court. On the other hand, we find no evidence that cases assigned to less experienced judges are more likely to be appealed. The coefficient on “deviating judicial preferences” indicate that a case is two percentage points more likely to be appealed if it is assigned to a judge who disagrees with a randomly drawn peer in 20% of the cases, compared to a judge who only disagrees with his or her judicial peers in 10% of the cases. The coefficient on mean case processing time indicate that if a case is assigned to a judge with a 10% longer leave-out mean case processing time the probability that the case is appealed is reduced by 0.25 percentage points. The latter result suggests that there might be a trade-off between speed and quality in judicial decision making.

In the third part of the paper we assess to what extent judges with a lower appeal rate

tend to be promoted to better positions. In order to classify judicial positions according to their attractiveness we distinguish between *large courts*, the eight largest District Courts located in Norway's main cities, and *small courts*. This distinction is meaningful since the number of applicants in to judicial jobs is much higher for the large courts. We find that judges with lower appeal rates, as measured by *past* cases, are *less* likely to obtain a judicial position at a new small court the next year compared to their peers with higher appeal rates. On the other hand, they seem to have a tendency of being more likely to obtain a position in a large court the following year, however this estimate is not statistically significant.

In the final part of the paper, we study the consequences of judges with high appeal tendencies. We find that judges with high appeal tendencies account for a significant share of appeals: If we replace all judges with more than average appeal tendencies among equally stringent judges with average judges with the same stringency we estimate that 4% of all appeals could be eliminated. In contrast, we find no evidence suggesting that the appeal tendency of the assigned judge causally affects defendants' recidivism or future labor market outcomes.

Our main contribution to the literature is providing what we think is the first evidence of the extent to which different judges have different tendencies of having their decisions appealed and reversed. Existing studies investigating such differences (e.g. Posner 2000) are in settings where cases are not randomly allocated between judges which means that we cannot rule out that differences in observed appeal rates or reversals are driven by differences in the types of cases decided. We also contribute to the literature seeking to measure the quality of judicial decision making. Most empirical studies investigating judicial quality either looks at the speed at which decisions are handed down or at judge citations (e.g. Choi and Gulati 2004; Ramseyer 2012; Palumbo *et al.* 2013), but a small subset looks at appeals and reversals (e.g. Posner 2000). By looking at whether fast judges tend to have higher appeal rate, we relate to Dimitrova-Grajzl *et al.* (2012) and Rosales-López (2008) who also investigate the possible existence of such a quantity-quality trade-off. However, these studies make no claim of cases being randomly assigned to judges. Finally, a growing

literature in law and economics has exploited the random assignment of cases to judges to estimate the causal impacts of trial court decisions on defendant outcomes (*e.g.* Di Tella and Schargrodsky 2013; Aizer and Doyle Jr 2015; Kling 2006; Dobbie *et al.* 2018) or to study biases in judge behavior (*e.g.* Shayo and Zussman 2011; Abrams *et al.* 2012). We contribute to this literature by using the same design to assess differences in appeal tendencies across judges.

2.2 Institutional Setting

We begin this section by reviewing key aspects of the Norwegian court system, documenting how cases are randomly assigned to judges and how court decisions can be appealed. Our setting is identical to Bhuller *et al.* (forthcoming), who utilized random assignment to district court judges to study the effects of incarceration on the defendant, and much of the discussion in Section 2.2.1 is taken from there. Notably, Bhuller *et al.* (forthcoming) did not consider the appeal process in Norway, which we describe in Section 2.2.2 below.

2.2.1 The Norwegian Court System

The court system in Norway consists of three levels: the district court, the court of appeals, and the supreme court. Similar systems with trial courts ('first instance') and appellate courts ('second or third instance') as in Norway exist in many other countries, including the U.S. where Courts of Appeals are empowered to hear appeals of decisions made in the district courts. In this paper, we focus on the first two levels of courts in Norway. The district courts and the courts of appeal process both criminal and civil cases. We consider only a subset of criminal cases that were tried in one of the 87 district courts in existence at one time or another during the period of our study, and that could have been appealed to one of the 6 courts of appeal in Norway.⁴

⁴As discussed below, the reasons for considering only a subset of criminal cases relates to data availability and that we were able to verify random assignment across judges in district courts for these cases.

Judges in District Courts. The largest district court is located in Oslo and has around 100 judges, while the smallest courts only have a few judges. As described in Bhuller *et al.* (forthcoming), there are two types of professional judges in district courts, regular judges and deputy judges. Regular judges are appointed civil servants, and can only be dismissed for malfeasance. One of the regular judges is appointed as chief judge to oversee the administration of the local court. Deputy judges, like regular judges, are also law school graduates, but are appointed to a court for a limited period of time which cannot exceed three years (five years in Oslo).⁵

Types of Criminal Cases. Criminal cases are classified into two broad types, confession and non-confession cases.⁶ In confession cases, the accused has confessed to the police/prosecutor before the case is assigned to a judge. The confession is entered into evidence, but the prosecution is not absolved of the duty to present a full case and the judge may still decide that the defendant is innocent.⁷ In practice, most confession cases are relatively straightforward. To save on time and costs, they are therefore heard by a single professional judge who decides on sentencing. Non-confession cases are heard by a panel of one professional and two lay judges, or in the case of extremely serious crimes, by two professional judges and three lay judges. The lay judges are individuals chosen from the general population to serve for a limited four year term. The professional judge presides over the case, while the lay judges participate on the questions of guilt and sentencing. As

⁵Deputy judges have a somewhat different caseload compared to regular judges, as discussed below. Not all deputy judges become regular judges, and those that do typically need several of years of experience in other legal settings before applying for and being appointed as a regular judge.

⁶The Norwegian criminal justice system does not have plea bargaining, so both confession and non-confession cases are settled by trial. By contrast, in the U.S. criminal defendants often know their assigned judge before deciding whether to plead guilty in exchange for a reduced sentence. Such pre-trial strategies make the interpretation of estimates from random-judge designs with plea bargaining harder to interpret (see, e.g., Dobbie *et al.*, 2018).

⁷These rules apply to most civil law systems, in contrast to common law systems where a majority of criminal cases are settled by confession and plea bargain rather than by a trial.

opposed to professional judges, lay judges hear only a few cases a year.⁸

Assignment of Cases in District Courts. In Norway, court guidelines prescribe that cases shall be assigned to judges according to the ‘principle of randomization’ (Bohn, 2000; NOU, 2002). The goal is to treat all cases ex-ante equally and prevent outsiders from influencing the process of the criminal justice system. In practice, cases are often assigned by the chief judge to other judges on a mechanical, rotating basis based on the date a case is received. Each time a new case arrives, it is assigned to the next judge on the list, with judges rotating between criminal and civil cases.⁹

As described in Bhuller *et al.* (forthcoming), there are some special instances where the assignment of cases does not follow the principle of randomization. These include cases involving juvenile offenders, cases with a statutory sentence length above six years, extremely serious cases which require two professional judges, and complex cases expected to take a longer time to process, all of which can be assigned to more experienced judges. Such cases are flagged in our data set. While all other cases are randomly assigned, some case types can only be assigned to regular judges, and deputy judges are assigned relatively more confession cases. This means that randomization occurs within judge type, but not necessarily across judge types. Therefore, to have a sample of *randomly assigned cases* to the *same pool of judges* we: (i) exclude the ‘special’ cases described above and (ii) focus on regular judges handling non-confession cases.

⁸Lay judges must satisfy certain requirements, such as not having a criminal record and not working in certain occupations (e.g., police officer). In a municipal district the pool of lay judges is usually between 30-60 individuals. Lay judges are partially compensated for days absent from work if not covered by their employer. We do not observe the identify of the lay judges in our data, but since they are randomly assigned to judges within a court, they should not create any bias in our estimates.

⁹This has been confirmed by servicemen in the Norwegian Courts Administration, the Oslo District Court, the Bergen District Court (the second largest court, after Oslo) and the Nedre Telemark District Court (a medium-sized court).

2.2.2 Appeals of District Court Decisions to the Courts of Appeal

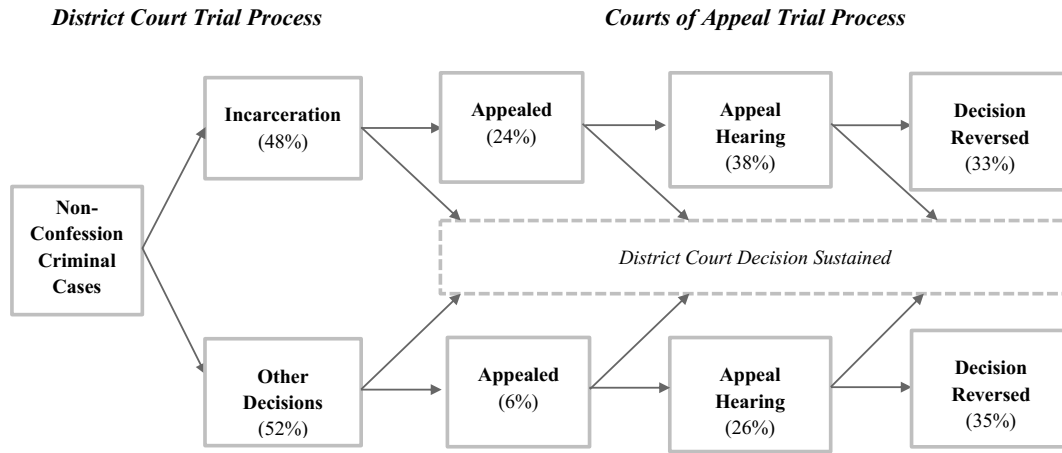
While the vast majority of court cases in Norway receive a final decision at the district court level, district court decisions can be appealed to the Courts of Appeal. An appeal can be filed by the prosecution, the defendant, or both, and may relate to district court guilt judgment, sentencing/penalty, application of the law, analysis of evidence, or procedural steps followed. In order for the appeal to be considered, it must be filed in writing within two weeks of the district court decision. A case decided in a district court can only be appealed to one of the 6 appeal courts in Norway, depending on the geographical jurisdiction pertaining to each appeal court.

The Courts of Appeal can either accept the appeal for a court hearing, or deny hearing. The appeals court can reject an appeal when the penalty is small, or if it finds that it is obvious that the appeal will not succeed.¹⁰ In the case of a rejection, the appeal court must issue a written ruling explaining the reasons to deny hearing. Appeals accepted for a hearing are decided by a panel of two professional and five lay judges.¹¹ Decisions in the appeal court are made by simple majority, with the exception of finding a defendant guilty which requires five votes. After concluding a court hearing, the appeal court may either decide to reverse (parts of) the district court decision, or reject the appeal and sustain the district court decision.

¹⁰An exception applies to cases with statutory sentence length above six years, which the Courts of Appeal are obliged to process for a full hearing if district court decision is appealed. As discussed in Section 2.2.1, these 'special' cases are excluded from our sample since such cases are not randomly assigned across all judges in district courts.

¹¹The jury ordinance was abolished on January 1, 2018, so cases appealed after this date are only decided by professional judges in the Courts of Appeal. All cases in our sample were decided or appealed before the new regulations went into affect. According to the prior regulation, appeal that do not concern the question of guilt can also have been decided by three professional judges, except cases with statutory sentence length above six years which could only be decided by two professional and three lay judges.

Figure 2.1: Case Trial and Appeal Process in Norway's Criminal Justice System.



Note: Sample consists of all non-confession criminal court cases processed in Norwegian district courts between 2005-2014 (N=159,020), including cases that are not necessarily randomly assigned and cases assigned that could have been assigned to deputy judges. Appeals and appeal outcomes are measured by the end of 2017 to avoid right-censoring. In our main analysis, we focus only on cases that are randomly assigned to regular judges.

Figure 2.1 charts how cases are processed in Norway's criminal justice system. The figure reports the number of cases at each stage in percentages across all non-confession criminal cases that were processed between 2005-2014, including the 'special cases' mentioned above and cases assigned to both regular and deputy judges. Our calculations show that almost 15% of decisions in district courts in such cases are appealed, with appeal rates rising to 24% for cases where the district court had ruled an incarceration decision and as low as 6% in cases with other district court decisions.¹² Two-thirds of the appeals are registered as 'comprehensive appeals', indicating that all aspects of district court decisions and proceedings are tried in the appeals court, while another one-fourth relate only to sentencing. Conditional on cases being appealed, only 35% of appeals are accepted for a court hearing by the Courts of Appeal. As shown in Figure 2.1, a higher fraction (38%) of appeals following an incarceration decision are accepted for an appeal hearing. Notably,

¹²The acquittal rate is at 5.5%, while the fraction of cases decided with a community service sentence is at 10%, probation at 17% and fine at 19.5%. Also see Bhuller *et al.* (forthcoming).

only in 33% of cases is a district court decision (partly or fully) reversed in the Courts of Appeal, conditional on being processed for a court hearing. Altogether, our calculations show that less than 2% of district court decisions later end up being reversed in an appeal court, even though a sizable fraction at 15% of cases are appealed.

2.3 Data Sources and Research Design

We begin this section by describing our data sources. We then provide tests of random assignment of cases to judges in district courts. Finally, we describe how to use this randomization to estimate the effects of a salient judge characteristic – judge appeal tendency – on subsequent appealing behavior and case and/or defendant outcomes.

2.3.1 Data and Sample Selection

Our analysis employs several data sources that we can link through unique identifiers for each case, defendant and judge. Information on the court cases comes from the Norwegian Courts Administration. The data set contains information for all court cases processed in district courts between 2005 and 2014. We observe the start and end dates of every trial, various case characteristics, the district court judgment, and identifiers for district courts. We can moreover follow appeals and appeal outcomes (appeal accepted/rejected for court hearing; decision reversed/sustained) in each of these cases in the Courts of Appeal up to the end of 2017. The three year post-decision window ensures that our appeal outcomes are not right-censored; by law appeals must be submitted within two weeks after the district court decision and only extremely exceptional cases can have a case processing time above three years in the Courts of Appeal.

We link this information with administrative data on all defendants involved in district court trials, which contain their complete labor market and criminal histories since 2000, including labor earnings, employment, registered charges, crime incidence information (type and date of crime), etc. We merge these data sets with administrative registers provided by Statistics Norway, using a rich longitudinal database that covers every resident from 1967

to 2016. For each year, it contains individual demographic information, including sex, age, and number of children, years of education, etc.

Our analytical sample consists of 74,818 non-confession cases randomly assigned to regular judges, excluding all confession cases and non-confession cases that could have been non-randomly assigned ('special' cases) or that were assigned to deputy judges; see description in Section 2.2.1.¹³ This yields a sample of randomly assigned cases to the same pool of judges. We further restricts the data set to judges who handle at least 50 randomly assigned cases between the years 2005 and 2014, which limits our sample to 71,069 cases. Since we will be including court by year of case registration fixed effects in all our estimates, we also limit the data set to courts which have at least two regular judges in a given year, dropping 2 district courts and limiting our sample to 70,109 cases. Dropping 4,193 non-resident aliens or refugees (6%), who cannot be linked to other data sources, we retain 65,916 cases in our baseline sample. This sample includes cases against 41,417 unique defendants decided by 539 district court judges in 85 district courts.

We now provide some summary statistics for defendants and crime types. Columns (5)-(6) in Table 2.1 provides means and standard deviations for all explanatory variables pertaining to defendant and case characteristics in our sample. Panel A shows that defendants are relatively likely to be young, single men with little education. Panel B shows they have high unemployment prior to the charge, with just above 33% of defendants working (earnings above 1 SGA) in the prior year. Serial offenders are common, with 45% of defendants having been charged for a different crime in the prior year. Panel C reports the fraction of cases by primary crime category. More than one fourth of cases involve violent crime, while property, economic, and drug crime each comprise a little more than 10 percent of crimes. Drunk driving, other traffic offenses, and miscellaneous crime make up

¹³Between 2005 and 2014, in total 264,611 criminal cases were processed in district courts in Norway. Among these, 159,020 (i.e., 60%) were non-confession and the rest were confession cases. Among the 159,020 non-confession cases, 50,273 (i.e., 32%) belonged to one of the 'special' cases that are not necessarily randomly assigned. Among the remaining 108,747 randomly assigned non-confession cases, 33,929 (i.e., 31%) were assigned to deputy judges. For reasons described in the text, we will thus focus on the resulting sample of 74,818 non-confession cases randomly assigned to regular judges in district courts.

the remainder.

2.3.2 Testing Random Assignment of Cases to Judges

Random assignment of cases to judges in district courts shall imply that characteristics of the assigned judge are unrelated to characteristics of the case or the defendant. As discussed in Section 2.2.1, assignment of cases to judges in the Norwegian court system is done on a mechanical, rotating basis based on the date a case is received at each district court. The actual assignment can therefore resemble random assignment only within groups of cases with the same court and case entry year. In order to verify the random assignment of such cases, we will thus perform balancing tests that condition on fully interacted court and year fixed effects.

Table 2.1 provides evidence consistent with cases in our sample being randomly assigned to judges. Column (1) regresses the probability that a district court decision is appealed (either by the defendant or the prosecutor) on a variety of variables measured before the court decision. It reveals that demographic, type of crime, and past criminal history variables are highly predictive of whether a defendant will appeal, with most of these being individually significant.

Using the collection of all randomly assigned cases to each judge – including some cases that were later dropped in our analytical sample – we constructed a measure of appeal tendency for each judge by leaving-out the current case observation. The appeal tendency of a judge thus equals the fraction of all other cases a judge has handled that later ended up being appealed to the Courts of Appeal.¹⁴ As we will discuss below, we consider a judge's appeal tendency to be a salient measure of quality in judicial decision making that can be related to future case or defendant outcomes.

¹⁴By leaving-out the current case observation in the construction of judge appeal tendency, we make sure that there is not mechanical relationship between appeal (outcomes) in the current case and the assigned judge's appeal tendency. With small samples, using the average judge appeal rate rather than a leave-out could be problematic for this reason. As discussed in Section 2.3.1, we restricted the data set to make sure that for each judge in our sample we have at least 50 cases that can be used to construct the leave-out judge appeal tendency. Judges entering our analytical sample have handled on average 201 cases each.

In column (3), we examine whether the same set of case and defendant characteristics that we used above can also predict this *leave-out* appeal tendency for the assigned judge. This is the same type of test that would be done to verify random assignment in a randomized controlled trial. We find that there is no statistically significant relationship between the judge appeal tendency variable and the various demographic, crime type and labor market variables. The estimates are all close to zero, with none of them being statistically significant at the 5% level. The variables are not jointly significant either (p-value=.704). This provides strong evidence that criminal court cases are randomly assigned to judges in our sample, conditional on fully interacted court and year fixed effects.

Table 2.1: Testing for Random Assignment of Criminal Cases to Judges.

	Pr(Decision Appealed)		Judge Leave-Out Appeal Tendency		Explanatory Variables:	
	(1)	(2)	(3)	(4)	(5)	(6)
	Coef.	Std.	Coef.	Std.	Mean	Std.
	Est.	Error	Est.	Error		Dev.
A. Demographics:						
Age	0.0026***	(0.0002)	-0.0000	(0.0000)	33.28	(11.59)
Female	-	(0.0032)	-0.0009*	(0.0006)	0.110	(0.313)
	0.0184***					
Foreign born	0.0053**	(0.0024)	0.0002	(0.0005)	0.160	(0.367)
Married, year t-1	0.0070	(0.0070)	-0.0013	(0.0009)	0.107	(0.309)
Number of children, year t-1	0.0046**	(0.0020)	-0.0002	(0.0002)	0.770	(1.241)
High school degree, year t-1	0.0076	(0.0050)	0.0005	(0.0007)	0.170	(0.376)
Some college, year t-1	0.0102	(0.0075)	0.0002	(0.0009)	0.049	(0.215)
Missing Xs	0.1878**	(0.0812)	0.0045	(0.0105)	0.054	(0.225)
B. Past Work and Criminal Hist.:						
Employed, year t-1	0.0026	(0.0043)	-0.0005	(0.0006)	0.333	(0.471)
Ever Employed, t-2 to t-5	0.0037	(0.0043)	-0.0004	(0.0006)	0.453	(0.498)
Charged, year t-1	-0.0049	(0.0037)	-0.0000	(0.0005)	0.453	(0.498)
Ever Charged, t-2 to t-5	-	(0.0038)	0.0001	(0.0006)	0.627	(0.484)
	0.0150***					
Incarcerated, year t-1	0.0151***	(0.0058)	0.0000	(0.0009)	0.133	(0.340)
Ever Incarcerated, t-2 to t-5	0.0186***	(0.0046)	0.0004	(0.0009)	0.283	(0.450)
C. Type of Crime:						
Violent crime	0.0493***	(0.0050)	0.0010*	(0.0006)	0.269	(0.444)
Property crime	-	(0.0051)	-0.0007	(0.0009)	0.131	(0.338)
	0.0448***					
Economic crime	-0.0084	(0.0070)	0.0007	(0.0009)	0.106	(0.309)
Drug related	-	(0.0052)	-0.0010	(0.0008)	0.133	(0.339)
	0.0364***					
Drunk driving	-0.0084	(0.0066)	-0.0004	(0.0009)	0.073	(0.260)
Other traffic	0.0054	(0.0074)	-0.0005	(0.0010)	0.071	(0.257)
F-statistic for Joint Test	41.36***		1.04		-	
[p-value]	[.000]		[.417]		-	
Dependent Mean	.151		.151		-	
Number of Cases	65,916		65,916		65,916	

Notes: Sample consists of non-confession criminal cases randomly assigned to regular judges in district courts 2005-2014. All estimations include controls for court x court entry year FEs. Reported F-statistic refers to a joint test of the null hypothesis for all variables. The omitted category for education is "Less than high school, year t-1" and the omitted category for type of crime is "Other crimes". Standard errors are two-way clustered at judge and defendant level. *p<0.1, **p<0.05, ***p<0.01.

2.3.3 Regression Model

We are interested in the causal effects of judge appeal tendency on subsequent appeals and other case and defendant outcomes. This can be captured by the following regression model:

$$Y_{ik} = \delta \bar{Z}_{j(ik)}^{-ik} + \theta X'_{ik} + \alpha_{ct} + \eta_{ik} \quad (2.1)$$

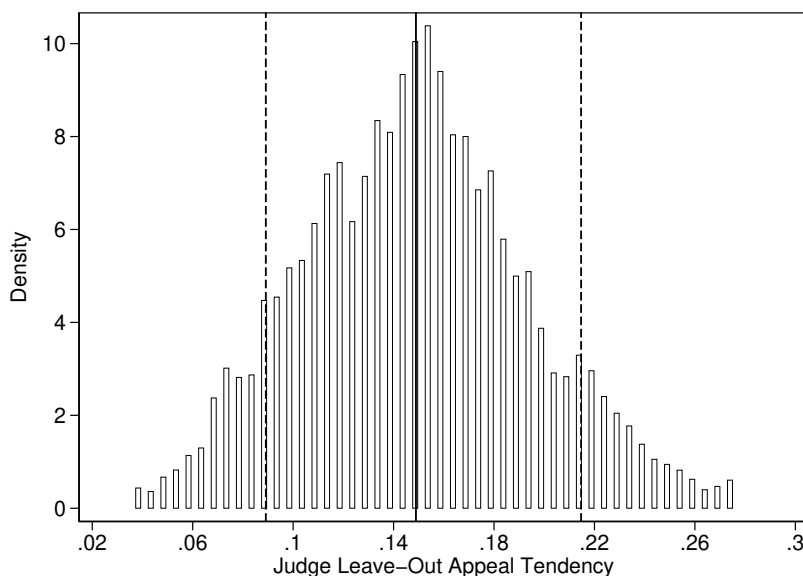
where δ is the parameter of interest, $\bar{Z}_{j(ik)}^{-ik}$ is the leave-out mean appeal tendency of judge j assigned to defendant i 's case k (superscript $-ik$ indicates that case observations for defendant i and case k are dropped when constructing this variable), X_{ik} is a vector of control variables capturing observed characteristics of defendant i and case k (that is, all variables listed in Table 2.1), α_{ct} are interacted court c and case year t fixed effects, and Y_{ik} is a dependent variable of interest measured at some later point in time (e.g., whether the trial court decision in case k was reversed in an appellate court, whether defendant i was charged with new crimes post-decision, etc).

The conditional random assignment of cases to judges – as verified in Table 2.1 – makes sure that conditional on α_{ct} , judge j 's appeal tendency $\bar{Z}_{j(ik)}^{-ik}$ is unrelated to (i) observed defendant and case characteristics X_{ik} and (ii) other unobserved determinants or correlates η_{ik} of Y_{ik} that may vary across defendants i and cases k . Due to (i), controlling for X_{ik} in (2.1) is not necessary to construct causal estimates of δ . In our empirical analysis, we will nonetheless show OLS estimates δ in (2.1) both with and without controls for X_{ik} . We also two-way cluster robust standard errors by judge j and defendant k in all our regressions to account for (i) assignment of cases is across judges and (ii) there can be multiple case observations of the same defendant.

A key to our research design is that not only are cases randomly assigned to judges, but also that judges differ in terms of their observed appeal tendencies. Figure 2.2 shows the distribution of appeal tendency across judges in our analytical sample, after residualizing on fully interacted court and year fixed effects as in our regression model. The plot illustrates striking differences in appeal rates across judges – while the average appeal rate is 15%,

judges with the lowest appeal rates are appealed in only 3% of the cases they have decided in a trial court, while judges with the highest appeal rates have around 27% of their decisions appealed.

Figure 2.2: *The Distribution of Judge Appeal Tendency.*



Notes: Figure 2.2 shows the distribution of appeal tendency across regular judges in district courts in Norway, handling randomly assigned non-confession criminal cases. The solid vertical line indicates the median in this distribution, while the dotted lines indicate the 10th and the 90th percentiles, respectively. The plotted values are mean-standardized residuals from regressions on court $\times$ court entry year interacted fixed effects. Judges with top and bottom 1% of appeal tendency are dropped in this plot.

Note that the variation in judge appeal tendency shown in Figure 2.2 does not necessarily represent systematic or true differences in judge appeal tendency across judges. In finite samples with a few cases handled by each judge, there can be differences in the characteristics of cases assigned to different judges even with random assignment of cases. If such case differences matter for appeal outcomes, then this will create variation in leave-out judge appeal tendencies, which may not reflect systematic differences across judges. Fortunately, we can easily assess the extent to which there are systematic differences in judge appeal tendencies with the data at hand and using the regression framework described above. In

particular, we can do this by estimating the following relationship:

$$Z_{ik} = \gamma \bar{Z}_{j(ik)}^{-ik} + \theta X'_{ik} + \alpha_{ct} + \epsilon_{ik} \quad (2.2)$$

where Z_{ik} is a binary indicator for whether the trial court decision made by judge j in case observation ik was later appealed to an appellate court. The OLS regression coefficient estimate of γ in (2.2) tells us to what extent the leave-out appeal rate of the assigned judge predicts the probability that a case will be appealed to an appellate court. As discussed in more detail in the Appendix Section B.1, under standard assumptions, this coefficient can also be interpreted as the share of the variance in observed (or leave-out) judge appeal tendency that can be attributed to systematic variation in appeal rates across judges as opposed to noise stemming from small sample variability in case characteristics across judges.

2.3.4 Judge Stringency and Judge Appeal Tendency

A number of previous empirical studies using random judge assignment designs exploit differences across judges in how often they tend to incarcerate defendants or differences in the sentence lengths they prescribe.¹⁵ In our context, one can imagine that defendants that receive harsher sentences are more likely to appeal. Differences in judge stringency could thus be a reason for observed differences in appeal tendencies across judges. In this section, we present how one can proceed (i) to test whether *all* of the judge-related variation in appeals is driven by differences in judge stringency, and (ii) to estimate the fraction of variation in appeal tendency across judges that can be attributed to differences in judge stringency.

Testing for Residual Variation in Judge Appeal Tendency. In order to test whether differences in appeal tendency across judges are driven only by differences in judge stringency,

¹⁵For instance, Kling (1999, 2006) exploits variation in sentence lengths across judges in the federal judicial system in California, while Bhuller *et al.* (forthcoming) exploit variation in incarceration stringency across judges using the same data set for Norway as we do (also see other references cited there for more recent work using similar designs).

let's first assume that the true appeal process is governed by the following equation:

$$Z_{ik} = bS_{ik} + cQ_{ik} + \xi_{ik}$$

where Z_{ik} equals one if case i with defendant k is appealed, S_{ik} is a vector of variables indicating the strictness of the sentence imposed by the judge, Q_{ik} is a vector with all other judge-related factors that might lead to a case being appealed, and ξ_{ik} is an iid error term, which captures all other non-judge factors. Factors entering Q_{ik} can include that the judge had wrongly applied the law, misunderstood the evidence, or made procedural mistakes, or a suspicion thereof. If all that matters for whether a case is appealed is the strictness of the sentence (prescribed by the judge) or idiosyncratic differences across defendants or cases (unrelated to the judge), then we have $c = 0$.

As earlier, let's denote the leave-out judge appeal tendency by $\bar{Z}_{j(ik)}^{-ik}$, and similarly denote the vector of leave-out judge stringencies by $\bar{S}_{j(ik)}^{-ik}$. Using OLS, we can relate both of these measures to an indicator for whether a case is appealed. To fit the setup described in Section 2.3.3, let's also imagine that we residualize each variable on court and case entry year fixed-effects and other covariates (for notational ease, we keep this implicit). The following proposition says that we can test for whether factors entering Q_{ik} matter for appeal outcomes based on OLS regression estimates.

Proposition 1. *Assume that*

$$Z_{ik} = \beta \bar{S}_{j(ik)}^{-ik} + \gamma \bar{Z}_{j(ik)}^{-ik} + \varepsilon_{ij} \quad (2.3)$$

with $\varepsilon_{ij} \perp (\bar{S}_{j(ik)}^{-ik}, \bar{Z}_{j(ik)}^{-ik})$. If the above assumptions hold and $c = 0$, then $\gamma = 0$.

The proof is provided in the Appendix B.2. The intuition behind this result is that if appeals are only determined by the strictness of a sentence, then the best predictor of whether a randomly drawn case is appealed is average judge stringency. The leave-out appeal tendency does not contain any information about judge stringency, once we control for leave-out judge stringency. Thus $\gamma = 0$. However, if factors entering Q_{ik} also cause appeals ($c \neq 0$), then leave-out appeal tendency will also predict appeal, beyond what

is predicted by leave-out stringency, and $\gamma \neq 0$. When we apply this test we will use conviction, incarceration, and the sentence length as measures of the strictness of a decision, S_{ik} .

The above proposition thus suggests that we can test for whether $c = 0$ by formally testing whether the coefficient $\gamma = 0$ in an OLS regression of equation (2.3). To verify that this is a reasonable test in small samples, we performed Monte Carlo simulations where we ran the above regression on simulated data with sample size matching our analytical sample under the assumption that $c = 0$. The parameters a and b were chosen to match observed appeal rates in our analytical sample. The results of this exercise are presented in the Appendix B.3. We plot the empirical distribution of t -values for $\hat{\gamma}$ in the Appendix Figure B.1. As expected the estimated t -values are centered around 0, however, their variance is higher than the standard normal under the assumption of $c = 0$. To adjust for this standard errors should be multiplied by a factor of 1.25, to give correct inference at the 5% level of significance according to the simulation exercise.

Variance Decomposition. According to the above proposition, we can test the null hypothesis that *all* of the judge-related variation in appeal rates are driven by differences in judge stringency, against the alternative hypothesis that there exists systematic variation in appeal rates due to other differences across judges. However, this test does not allow us to say *how much* of the variation in appeal rates is due to differences in judge stringency, and how much of the variation is due to other differences across judges. In order to investigate this, let's denote the true (large sample) appeal tendency and stringency of judge j by $\bar{Z}_j$ and $\bar{S}_j$, respectively, and assume that

$$\bar{Z}_j = \mu \bar{S}_j + v_j$$

where v_j is an iid error term. If $\mu > 0$ this means that more stringent judges also tend to have a higher appeal tendency. Under this assumption the variance in appeal tendency across judges can be decomposed into the variance due to differences in judge stringency,

and the remaining variance as follows

$$\text{Var}(\bar{Z}_j) = \mu^2 \text{Var}(\bar{S}_j) + \text{Var}(v_j)$$

We already have an estimate of $\text{Var}(\bar{Z}_j)$ from Section 2.3.3. The following proposition suggests how to estimate the contribution of judge stringencies to the overall variance in judge appeal rates.

Proposition 2. *Under the above assumption*

$$\text{plim} \frac{\text{Cov}\left(Z_{j(ik)}, \bar{S}_{j(ik)}^{-ik}\right)^2}{\text{Cov}\left(S_{j(ik)}, \bar{S}_{j(ik)}^{-ik}\right)} = \mu^2 \text{Var}(\bar{S}_j)$$

The proof as well as a generalization to multi-dimensional stringency is provided in the Appendix B.2.

2.3.5 Other Judge Characteristics

Finally, it is natural to ask whether there are other reasons why some judges are more likely to have higher appeal tendencies than others besides differences in their stringency. While we do not observe personal characteristics of judges in our data for privacy reasons, we can measure a number of other salient judge characteristics using the same data set, such as a measure of each judge's tendency to judge cases at odds compared to their judicial peers (i.e., *judge disagreement*; see more on this below), leave-out mean case processing time (i.e., *judicial efficiency*), and how long each judge have been in judicial service (i.e., *judicial experience*) or worked at a particular court (i.e., *court experience*). In order to gauge the importance of these judge characteristics for appeal outcomes, we will also provide results from the following regressions:

$$Z_{ik} = \beta \bar{S}_{j(ik)}^{-ik} + \varphi \bar{J}_{j(ik)}^{-ik} + \theta X'_{ik} + \alpha_{ct} + \varepsilon_{ik} \quad (2.4)$$

where $\bar{J}_{j(ik)}^{-ik}$ is a vector of the other observed characteristics of judge j described above and the coefficient vector φ captures how these characteristics affect the probability that a case is appealed, conditional on judge stringency measures. As earlier, we control for a vector of

defendant and case characteristics X_{ik} and a full set of interacted court and case entry year fixed effects.

Measuring Judge Disagreement. We present here an approach on how we construct a measure of whether a particular judge tends to arrive at case decisions which systematically deviate from the decisions of his or her judicial peers in cases that are observationally similar. Let's consider that judges can have different preferences regarding whether or not to impose an incarceration sentence based on the same set of characteristics of a case. In particular, let's assume that judge j incarcerates iff

$$\lambda_j X'_{ik} + \kappa_{ikj} \geq 0$$

where X'_{ik} is a vector of case and defendant characteristics and κ_{ikj} is an iid extreme value distributed error term across ik . Deviance in judicial preferences across judges is allowed here by letting the coefficient vector λ_j be judge-specific.¹⁶ To fix ideas, let's consider that X'_{ik} is composed of the variables “drug-related crime” and “low income defendant”. Judges can then deviate in their decision making, for instance, by having a higher than average coefficient on “drug-related crime” (being relatively strict on drug-related crimes) or by having a lower coefficient on “low income defendant” (being relatively lenient on low income defendants). Given the distributional assumption on the error terms, one can estimate the coefficient vector λ_j by estimating a standard logit choice model separately for each judge.

As detailed in Appendix B.4, using the estimated coefficient vector $\hat{\lambda}_j$ we construct a summary measure of how each judge is expected to deviate from his or her judicial peers. Specifically, for each pair of judges working in the same court we can calculate the probability that this pair of judges will decide a randomly assigned case differently based on

¹⁶In a related approach, Fischman (2013) introduces a measure of judicial inconsistency and provides conditions under which this measure can be non-parametrically bounded. Our judge disagreement measure corresponds closely to the lower bound of judicial inconsistency in Fischman (2013), which assumes monotonicity in judge behavior. By contrast, our approach allows judge behavior to depend on covariates and assumes only conditional monotonicity, while imposing separability between covariates and errors.

the estimated coefficient vector $\hat{\lambda}_j$. Performing this exercise for all possible pairs of judges in each court and then averaging over the probability differences for each judge, we can construct a summary measure of judge disagreement. For this exercise, we use type of crime and defendant’s education, gender, past employment and criminal history as covariates, and use regularization to select the set of variables that are most predictive of incarceration decisions for each judge to avoid overfitting.

2.4 Judge Appeal Tendency and Appeal Behavior

In this section, we present our main findings, showing that (i) the leave-out judge appeal tendency for the assigned judge is highly predictive of whether a case is appealed and whether a case decision is reversed in an appellate court, (ii) a sizable fraction of the variation in leave-out appeal rates can be attributed to systematic differences in appeal tendencies across judges, and (iii) the relationship between leave-out judge appeal tendency and appeal outcomes is robust to controlling for judge stringency with respect to conviction, incarceration, and sentence length, and a substantial part of the variation in systematic judge appeal tendency cannot be attributed to differences in judge stringency.

2.4.1 Main Results: The Effects of Judge Appeal Tendency on Appeal Outcomes

We start by documenting the relationship between leave-out judge appeal tendency of the assigned judge and appeal outcomes for the cases handled by this judge. Figure 2.3a shows the identifying variation in our analytical sample and provides a graphical representation of this relationship. The histogram in the background shows the distribution of leave-out appeal tendency, after residualizing on fully interacted court and year fixed effects, as also illustrated in Section 2.3.3. The plot also illustrates the probability that a case is appealed as a function of leave-out appeal tendency of the trial court judge the case is assigned to, estimated using a local linear regression. The likelihood that a trial case decision is appealed is estimated to be monotonically increasing in the appeal tendency of the assigned trial court judge.

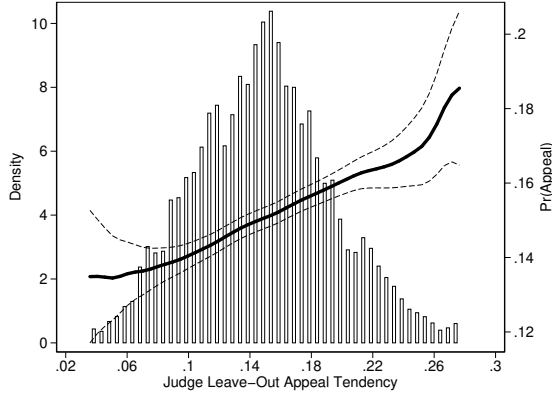
Table 2.2: *The Effects of Judge Appeal Tendency on Appeal Outcomes.*

	Pr(Decision Appealed)	Pr(Appeal Processed)	Pr(Decision Reversed)
	(1)	(2)	(3)
A. Court x Year of Court Case Registration Interacted Fixed Effects			
Appeal Tendency	0.218***	0.082***	0.078***
	(0.051)	(0.023)	(0.018)
B. Add Controls for Defendants' Demographics			
Appeal Tendency	0.224***	0.085***	0.079***
	(0.050)	(0.023)	(0.017)
C. Add Controls for Defendants' Demographics and Past Work & Criminal History			
Appeal Tendency	0.225***	0.085***	0.079***
	(0.050)	(0.023)	(0.018)
D. Add Controls for Defendants' Demographics, Past Work & Criminal History, and Type of Crime			
Appeal Tendency	0.217***	0.082***	0.078***
	(0.050)	(0.023)	(0.017)
Controls for α_{ct}	✓	✓	✓
Dependent Mean	.151	.042	.016
Number of Cases	65,916	65,916	65,916

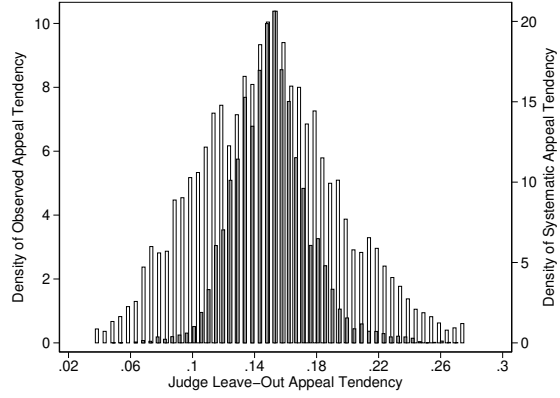
Note: Sample consists of non-confession criminal cases processed 2005-2014. All estimations include controls for court x court entry year FEs. Standard errors are two-way clustered at judge and defendant level. *p<0.1, **p<0.05, ***p<0.01.

Figure 2.3: *The Distribution of Leave-Out and Systematic Judge Appeal Tendency.*

(a) *Pr(Decision Appealed) on Leave-Out Judge Appeal Tendency*



(b) *Systematic Judge Appeal Tendency*



Notes: Plot 2.3a shows (i) the distribution of observed appeal tendency (left y-axis) across regular judges in district courts in Norway, handling randomly assigned non-confession criminal cases, and (ii) the probability of a case decision being appealed (along the right y-axis) against leave-out judge appeal tendency (along the x-axis), estimated using a local linear regression. Plot 2.3b shows (i) the distribution of observed appeal tendency (left y-axis) across regular judges in district courts in Norway, and (ii) the distribution of systematic appeal tendency (right y-axis) constructed using the estimates reported in Table 2.2, column (1). The plotted values are mean-standardized residuals from regressions on court $\times$ court entry year interacted fixed effects. Judges with top and bottom 1% of appeal tendency are dropped in each plot.

To translate the graphical evidence to our regression framework, we estimate equation (2.2), controlling for fully interacted court and court entry year fixed-effects. The results are provided in Table 2.2, panel A, column (1). The point estimate of 0.0218 indicates that a case has 2.18 percentage points increased probability of being appealed when assigned to a judge with 10 percent points higher leave-out appeal tendency. The standard errors are heteroskedastic- and cluster-robust, and the point estimate is statistically significant at 1% significance level and sizable compared to the mean appeal rate of 15.1% in our sample. In panels B-D, we further control for a vector of defendants' demographics, their past work and criminal history, and type of crime (see Table 2.1 for a complete list of variables included). As expected given the random assignment of criminal cases within court and entry year cells, inclusion of additional controls for defendant and case characteristics does not influence the point estimates of judge appeal tendency on whether a case is appealed.

Systematic Variation in Judge Appeal Tendency. As discussed earlier, the variation in judge appeal tendency shown in Figure 2.3a does not necessarily represent systematic or true differences in judge appeal tendency across judges. In finite samples with a few cases handled by each judge, there can be differences in the characteristics of cases assigned to different judges even with random assignment of cases. If such case differences matter for appeal outcomes, then this will create variation in leave-out judge appeal tendency, which may not reflect systematic differences across judges. Using the point estimates reported in Table 2.2, column (1), and building on insights developed in Appendix B.1, we next provide an estimate of systematic variance in judge appeal tendency below that removes variation in appeal tendency due to noise.

As shown in Table 2.3, we calculate that the standard deviation of the raw measure of leave-out appeal tendency in our analytical sample is at 0.0746. Before performing the adjustment for finite sample noise in this measure, we residualize on court and case entry year fixed-effects (i.e., the level of randomization of cases to judges). This is done to fit the setup described in Section 2.3.3 and to isolate the variation in judge appeal tendency in our regression model. Note that the variation in judge appeal tendency shown in Figure 2.3a is also residualized in the same manner. Our calculations reveal that the standard deviation of the residualized leave-out appeal tendency is at 0.0517, which implies that 48% of the variance in raw leave-out appeal tendency can be attributed to differences *across* courts and case entry year cells, while the remaining fraction pertains to differences in appeal tendencies across judges *within* court-year cells.

Next, using the approach described in Appendix B.1, we perform a noise adjustment to isolate the systematic part of the variation in appeal tendencies across judges. Our calculations reveal that the standard deviation of the systematic leave-out appeal tendency is at 0.0241, which implies that 10.4% of the variance in raw leave-out appeal tendency and around 20% of the variance in residualized leave-out appeal tendency can be attributed to systematic across-judge differences in appeal tendencies. For illustration, we overlay the histograms of residualized leave-out appeal tendency (gray-colored bars) and systematic

appeal tendency (dark-colored bars) in Figure 2.3b. Focusing on the systematic part of appeal tendency differences, this plot reveals that the vast majority of judges in our analytical sample have a appeal rate between 8% and 22% – with a mean appeal tendency across judges at 15%.

Table 2.3: *Decomposition of the Variance of Raw Judge Appeal Tendency.*

		Standard Deviation	%Variance %Explained	%Variance %Remaining
		(1)	(2)	(3)
(1)	Leave-Out Appeal Tendency <i>raw measure</i>	0.0746	-	100%
(2)	Residualized Appeal Tendency <i>residualize on α_{ct}</i>	0.0517	48%	52%
(3)	Systematic Appeal Tendency <i>perform noise adjustment</i>	0.0241	89.6%	10.4%

Note: Sample consists of non-confession criminal cases processed 2005-2014.

Appeal Court Hearings and Decision Reversals. As discussed in Section 2.2.2, once a case is appealed to the Courts of Appeal, an appeal court judge can either deny the appeal a court hearing or accept hearing, and in the latter case, the trial court decision can either be reversed or sustained. To assess whether the leave-out judge appeal tendency of the assigned trial court judge also impacts the later appeal outcomes in appellate courts, we estimate equation (2.1) on two additional indicator variables. The first of these outcomes equals one only if the appeal is processed for a hearing in the Courts of Appeal, while the second outcome equals one only if the trial court decision is (partly or fully) reversed. The estimates of leave-out judge appeal tendency on these later appeal outcomes are provided in Table 2.2, columns (2)-(3). These estimates indicate that appeal tendency of a trial court also significantly affect the likelihood that a case is processed for an appeal hearing and

whether the trial court decision is later reversed. As earlier, these estimates are robust to inclusion of controls for defendant and case characteristics.

Next, we consider whether cases assigned to trial court judges with a higher appeal tendency also have higher probability of decision reversal, conditional on being appealed. In other words, do estimates in Table 2.2, column (3), simply reflect that judges with higher appeal tendency also have higher probability of receiving an appeal (and a fixed share of such case decisions gets reversed) or do such judges also have a higher conditional probability of a decision reversal, for instance, if the quality of their judicial decision making is worse than other cases that are appealed? Since we cannot randomize appeal behavior and perform causal comparisons across conditional outcomes in our setting, decomposing these two sources is not straightforward. The point estimates in Table 2.2 are nonetheless informative. To see this, note that the overall fraction of cases appealed in our sample that received a decision reversal equals 10.6% ($\simeq 0.016/0.151$). Using the point estimates in panel D, columns (1) and (3), we calculate the fraction of cases with a decision reversal conditional on being appealed equals 13.8% ($\simeq (0.016 + 0.078 \times 0.1) / (0.151 + 0.217 \times 0.1)$) when such cases are assigned to judges with a 10 percent point higher appeal tendency, indicating that such judges have around 30% higher conditional probability of decision reversal.

2.4.2 The Role of Judge Stringency

Using the framework presented in Section 2.3.4, we here provide results from (i) a test of whether *all* of the judge-related variation in appeals is driven by differences in judge stringency, and (ii) a decomposition of the variance in systematic appeal tendency differences across judges that can be attributed to differences in judge stringency and other judge-related factors.

To perform (i), we estimate equation (2.3) including as explanatory variables leave-out measures of appeal tendency, incarceration stringency, conviction stringency and sentence length stringency for the assigned judge, and further also condition on interacted court

and case entry year fixed-effects (α_{ct}) and a vector of defendant and case characteristics (X'_{ik}). The results from this exercise are presented in Table (2.4), column (1). We find that judge incarceration stringency has a significant effect on the probability that a case decision is appealed, while we are unable to reject the null hypothesis for the other two measures of judge stringency. Importantly, we find that the coefficient on leave-out judge appeal tendency is significant and positive. Thus, based on this evidence and proposition (1), we can conclude that not all of the judge-related variation in appeals is driven by differences in judge stringency.¹⁷

Table 2.4: *The Effects of Judge Stringency on Appeal Outcomes.*

	Pr(Decision Appealed)	Pr(Appeal Processed)	Pr(Decision Reversed)
	(1)	(2)	(3)
Appeal Tendency	0.162*** (0.053)	0.061** (0.024)	0.070*** (0.018)
Incarceration Stringency	0.198*** (0.046)	0.052*** (0.025)	0.060*** (0.016)
Sentence Length Stringency	-0.031 (0.020)	-0.023** (0.010)	-0.017** (0.008)
Conviction Stringency	0.075 (0.123)	-0.046 (0.056)	-0.070** (0.035)
Controls for α_{ct} and X'_{ik}	✓	✓	✓
Dependent Mean	.151	.042	.016
Number of Cases	65,916	65,916	65,916

Note: Sample consists of non-confession criminal cases processed 2005-2014. All estimations include controls for court x court entry year FEs and all defendants and case characteristics listed in Table 2.1. Standard errors are two-way clustered at judge and defendant level. *p<0.1, **p<0.05, ***p<0.01.

¹⁷This conclusion remains unchanged if we multiply the standard errors by 1.25, as suggested by the Monte Carlo simulations we performed in Appendix (B.3).

For completeness, we also estimated the same specification as discussed above on indicators for whether a case was processed for appeal hearing and whether the trial case decision was later reversed in the Courts of Appeal. These estimates are presented in Table 2.4, columns (2)-(3). We again find that the effects of leave-out judge appeal tendency on the two later appeal outcomes remain statistically significant and positive. Our previous conclusion regarding judges with higher appeal tendency having a higher conditional probability of decision reversal remains robust to inclusion of judge stringency measures in the estimation.

Next, we move on to the variance decomposition analysis. Using proposition (2), we perform a decomposition of the variance in systematic appeal tendency differences across judges that can be attributed to differences in judge stringency and to other judge-related factors. The results from this decomposition are presented in Table 2.5. In row (1), we repeat our previous estimates of the systematic variation in leave-out appeal tendency from Table 2.3, row (3). Our further calculations reveal that the standard deviation of the systematic appeal tendency declines to 0.0163 once we remove the component that can be attributed to differences in judge incarceration stringency. This estimate indicates that 54% of the variance in systematic appeal tendency can be attributed to judge incarceration stringency. In rows (3)-(4), we further remove variation due to differences in judge conviction and sentence length stringencies, and find that 39% of the variance in systematic appeal tendency cannot be explained jointly by the three judge stringency measures. Thus, we conclude that a substantial fraction of appeal tendency differences across judges are likely to be related to other non-stringency judge factors.

Table 2.5: *Decomposition of the Variance of Systematic Judge Appeal Tendency.*

	Standard Deviation	%Variance %Explained	%Variance %Remaining
	(1)	(2)	(3)
(1) Systematic Appeal Tendency	0.0241	-	100%
(2) Remove Incarceration Stringency	0.0163	54%	46%
(3) Remove Sentence Length Stringency	0.0152	60%	40%
(4) Remove Conviction Stringency	0.0150	61%	39%

Note: Sample consists of non-confession criminal cases processed 2005-2014.

2.4.3 The Role of Other Judge Characteristics

Why do certain judges seem to have a higher appeal rate even controlling for the harshness of their sentences? There might be several reasons. This could happen if some judges apply the law differently than others, or if some judges are more likely to misinterpret the evidence or make procedural errors. This again could be caused by a lack of experience or knowledge, by judges not investing enough time or effort into considering each case, or simply by deviant judge preferences. In this section we try to answer which of these factors might explain why some judges have their decisions appealed at a higher rate than others.

In order to do this, we construct several additional judge characteristics using our data set that are intended to capture different attributes of a judge which might cause appeals due to any of the above-mentioned reasons. We then use the specification described in Section 2.3.5 to test whether a case is more likely to be appealed if it is randomly assigned to a judge with a certain characteristic. The variables we construct are the following. As a

proxy for experience we use *judicial tenure*, which is the log number of months the judge is observed in our data. We complement this measure with *court tenure*, the log number of months the judge has been at the current court. To measure the speed at which different judges make decisions we use *case processing time*, the leave-out mean log number of days between filing and decision for cases decided by the judge. Finally, we construct the variable *disagreement tendency* as described in Section 2.3.5 to proxy for the extent to which the preferences of the judge deviate from his or her judicial peers.

Table 2.6: *The Effects of Other Judge Characteristics on Appeal Outcomes.*

	Pr(Decision Appealed)	Pr(Appeal Processed)	Pr(Decision Reversed)
	(1)	(2)	(3)
Incarceration Stringency	0.288*** (0.055)	0.126*** (0.027)	0.078*** (0.019)
Sentence Length Stringency	-0.034 (0.023)	-0.022** (0.010)	-0.017* (0.009)
Conviction Stringency	0.074 (0.133)	-0.016 (0.059)	-0.042 (0.041)
Disagreement Tendency	0.249** (0.121)	0.082 (0.066)	-0.002 (0.044)
Case Processing Time (in logs)	-0.025** (0.012)	-0.008 (0.006)	-0.001 (0.004)
Judicial Tenure (in logs)	-0.003 (0.003)	-0.001 (0.002)	-0.001 (0.001)
Court Tenure (in logs)	0.004** (0.002)	0.002* (0.001)	0.001** (0.001)
Controls for X'_{ik} and α_{ct}	✓	✓	✓
Controls for Judge Stringency	✓	✓	✓
Dependent Mean	.150	.042	.016
Number of Cases	64,711	64,711	64,711

Note: Sample consists of non-confession criminal cases processed 2005-2014. All estimations include controls for court x court entry year FEs, all defendants and case characteristics listed in Table 2.1 and leave-out judge stringency. Standard errors are two-way clustered at judge and defendant level. *p<0.1, **p<0.05, ***p<0.01.

The results are presented in Table 2.6, column (1). As discussed earlier, the estimates in the first row indicate that judge incarceration stringency positively affects the probability that a case is appealed, while the coefficients on sentence length stringency (second row)

and conviction stringency (third row) are statistically insignificant.

Moving on to judge disagreement tendency in the fourth row, we find that a case is more likely to be appealed if it is assigned to a more disagreeing judge. The interpretation of this coefficient is that a case is 2 percentage points more likely to be appealed if it is assigned to a judge who disagrees with a randomly drawn peer in 20% of the cases, compared to a judge who only disagrees with his or her judicial peers in 10% of the cases. This evidence is consistent with the notion that judges with more deviant preferences are more likely to face appeals.

In the fifth row, we report the estimates on leave-out mean case processing time. Our estimates indicate that if a case is randomly assigned to a slower judge (i.e., longer processing time), the case is actually less likely to be appealed. If the case is assigned to a judge with a 10% longer leave-out mean case processing time the probability that the case is appealed is reduced by 0.25 percentage points. This evidence is consistent with fast judges committing mistakes and to a possible quality-quantity trade-off in judicial decision-making.

Next, we consider overall judicial tenure and court-specific tenure. While the coefficient on judicial tenure is statistically insignificant, the point estimate is negative, consistent with appeals being less likely if cases are assigned to more experienced judges. By contrast, the coefficient on court tenure is positive and statistically significant at the 5 percent level, with the point estimate indicating that a case is 0.04 percentage points more likely to be appealed if assigned to a judge who has been 10% longer in that particular court. This result could be due to (i) dynamic selection of judges, i.e., judges with higher appeal rates more less likely to stay for a long time in the same court, and/or due to (ii) judges' career incentives, i.e., judges who tend to stay a long time in a given court are, as showed in Section 2.4.4, more likely to continue in the same court, and might thus have fewer career concerns.

For completeness, we also estimated the same specification as discussed above on indicators for whether a case was processed for appeal hearing and whether the trial case decision was later reversed in the Courts of Appeal. These estimates are presented in Table 2.6, columns (2)-(3). With the exception of court tenure and incarceration stringency, we find

no statistically detectable differences across judges with different characteristics for these two later appeal outcomes at the 5 percent level. However, the coefficients on most judge characteristics show the same sign across the three appeal outcomes reported in columns (1)-(3).

2.4.4 Judge Appeal Tendencies and Career Outcomes

In the hiring process of regular judges (who all have past experience as deputy judges), it is plausible to assume that a low prior appeal tendency is regarded as a desirable attribute of the job applicant. Thus, one might expect that judges with a lower past appeal tendency would have an easier time to obtain a position in a new court. In this section we investigate whether this seems to be the case. To do this, we calculate the judge appeal tendency only based on *past* cases decided by the judge and run our main specification with the outcomes being whether the judge in the next year (i) stays in the same court, (ii) works at a different trial court, or (iii) is not observed in our data. The latter might indicate retirement, judges finding jobs outside the judiciary, or appointments to an appeals court.¹⁸

To further distinguish job transfers due to promotion and due to other reasons, we split the outcome of working in a different trial court into two mutually exclusive outcomes. We classify transfers to any one of the eight largest district courts which are located in Norway's largest cities as being a "promotion", while the remaining job transfers are simply labeled as a "transfer". Job openings in the larger courts tend to have much more applicants than do openings in the smaller courts.¹⁹ However, all transfers require the same application process.²⁰ To isolate the role of appeal tendency for career outcomes from other salient judge characteristics, we control for judge stringency, as well as judicial tenure and court

¹⁸Unfortunately, our data do not allow us to tell whether a trial court judge is nominated to an appeals court.

¹⁹The average judge position had only 9 applicants in 2017, whereas the position with the most applicants had 58 (source: <https://www.domstol.no/globalassets/upload/da/domstol.no/om-domstolene/innstillingsradet/arsmeldinger/arsmelding-2017.pdf>)

²⁰Judges in Norway are appointed by a judicial council composed of three judges, two lawyers, where one has to be a public employee, and two non-lawyers. The council members are themselves appointed by the executive and serve for a four year term.

tenure, in this analysis.

Table 2.7: *Judge Appeal Tendency, Judicial Tenure and Career Progression.*

	Pr(Staying in Year t+1)	Pr(Transfer in Year t+1)	Pr(Promotion in Year t+1)	Pr(Exit in Year t+1)
	(1)	(2)	(3)	(4)
Past Appeal Tendency	-0.372 (0.249)	0.542** (0.233)	-0.174 (0.135)	-0.005 (0.103)
Judicial Tenure (in logs)	0.085*** (0.010)	-0.016* (0.008)	-0.034*** (0.006)	-0.036*** (0.003)
Court Tenure (in logs)	0.027*** (0.009)	-0.004 (0.008)	-0.006* (0.003)	-0.017*** (0.003)
Controls for X'_{ik} and α_{ct}	✓	✓	✓	✓
Controls for Judge Stringency	✓	✓	✓	✓
Dependent Mean	.770	.133	.057	.040
Number of Cases	64,693	64,693	64,693	64,693

Note: Sample consists of non-confession criminal cases processed 2005-2014. All estimations include controls for court x court entry year FEs, all defendants and case characteristics listed in Table 2.1 and leave-out judge stringency. Standard errors are two-way clustered at judge and defendant level. *p<0.1, **p<0.05, ***p<0.01.

The results are presented in Table 2.7. The first row shows the effects of past appeal tendency on judge career outcomes. The only statistically significant coefficient is for the “transfer” outcome in column (2). Judges with higher past appeal rates have a *higher* probability of working in another smaller trial court in the subsequent year. The coefficient indicates that a judge with a one percentage point higher past appeal tendency is 0.5 percentage points more likely to be working at another smaller trial court the next year. This is a large point estimate, given that the baseline probability that a judge works in another smaller trial court the next year is 13.3%. At first glance, our evidence can be seen as being inconsistent with the notion that judges with higher appeal tendency have a more

difficult time obtaining new appointments. The result however could also be explained by self-selection; judges with higher appeal rates could be more willing to start working at another court, and smaller courts usually tend to have fewer applicants, and thus such judges might more easily find jobs at such places. The results in column (3) moreover suggest that judges with a *lower* appeal tendency do have a higher likelihood of getting the more sought after jobs in the larger courts. However, the coefficient is not statistically different from zero.

2.5 Judge Appeal Tendency, Judicial Costs and Defendant Outcomes

In this section, we investigate the *consequences* of case assignment to judges with higher than average appeal tendency. First, we use data on judicial accounts to assess the direct judicial costs associated with assignment to higher appeal tendency judges. Second, we provide a back-of-the-envelope calculation of the share of appeals that could be eliminated if judges with more than average appeal tendencies among judges with the same level of stringency, were replaced by judges this average appeal tendency. Finally, linking our data set to future outcomes for defendants, we assess the impacts on defendants' recidivism and labor market outcomes.

2.5.1 The Effects of Judge Appeal Tendency on Judicial Costs

To assess the judicial costs associated with case assignment to high appeal tendency judges, we use data on judicial accounts for each District Court and Courts of Appeal for the year of 2017 provided by the Norwegian Courts Administration. For each court we calculate the average judicial cost of processing a court case per day as follows. First, we calculate the total number of days between case filing and decision across all cases (including civil cases) processed in each court in 2017. Next, we divide the total annual court budget on the aggregate number of case days across all processed cases. Using this average per-case day

cost estimate, we calculate the direct judicial costs associated with each court case. Finally, we use this per-case cost estimate as an outcome in our regression model.

The estimates resulting from this exercise are provided in Table 2.8. We start by showing the estimates of appeal tendency on case processing time in the Court of Appeal (column 1), total case processing time (column 2), judicial costs in the Court of Appeal (column 3), and total judicial costs (column 4). The results in panel A show that being assigned to a judge with a 10 percentage points higher appeal tendency leads to on average 1.4 days of case processing time and to an increase in direct judicial cost by NOK 17,071 ($\approx$ 2000 USD) in the Courts of Appeal. The point estimates remain similar once we take into account case processing time and direct judicial costs imposed in the District Court.

Table 2.8: *The Effects of Judge Appeal Tendency on Judicial Costs.*

	Processing Time in the Court of Appeal (days)	Processing Time in the District Court and the Court of Appeal (days)	Judicial Costs in the Court of Appeal (2017 NOK)	Judicial Costs in the District Court and the Court of Appeal (2017 NOK)
	(1)	(2)	(3)	(4)
A. Judge Appeal Tendency				
Appeal Tendency	14.65*** (4.96)	14.12 (9.09)	17,071*** (6,212)	17,489** (8,869)
Controls for X'_{ik} and α_{ct}	✓	✓	✓	✓
B. Judge Appeal Tendency – Controlling for Judge Stringency and Disagreement				
Appeal Tendency	10.82** (5.15)	9.90 (9.85)	12,875** (6,474)	13,601 (9,310)
Controls for X'_{ik} and α_{ct}	✓	✓	✓	✓
Controls for Judge Stringency	✓	✓	✓	✓
Controls for Judge Disagreement	✓	✓	✓	✓
Dependent Mean	8.13	111.7	9,806	93,310
Dependent Mean (conditional on appeal)	54.01	166.8	65,122	159,088
Number of Cases	65,916	65,916	65,916	65,916

Note: Sample consists of non-confession criminal cases processed 2005-2014. All estimations include controls for court x court entry year FEs and all defendants and case characteristics listed in Table 2.1. Panel B also add controls for assigned judge's leave-out incarceration stringency, sentence length stringency, conviction stringency, and disagreement tendency. Standard errors are two-way clustered at judge and defendant level. *p<0.1, **p<0.05, ***p<0

2.5.2 The Share of Appeals Due to Judges with High Appeal Tendencies

How much could appeals be reduced if we replace judges with an appeal tendencies higher than the average among judges with the same level of stringency with this average? To

provide a back-of-the-envelope calculation of this reduction, we assume, in addition to the assumptions behind the variance decomposition exercise in Section 2.3.4, that judge appeal tendency is normally distributed conditional on judge stringency. The reduction in appeals as a share of all cases decided is then equal to²¹

$$\sqrt{\frac{\text{Var}(v_j)}{2\pi}}$$

where $\text{Var}(v_j)$ is the variance in appeal tendencies unexplained by judge stringencies. Using the estimate from Section 2.4.2 we thus get that the share of appeals would be reduced by $0.015/\sqrt{2\pi} = 0.006$. Given that the average appeal rate is 0.151 this corresponds to a reduction of 4% in the total number of appeals.

2.5.3 The Effects of Judge Appeal Tendency on Defendants' Future Outcomes

To assess the effects of case assignment to high appeal tendency judges on defendants' future outcomes, we link our data set to future recidivism, employment and earnings history over the five first years post decision, as in Bhuller *et al.* (forthcoming). Since we only have available future defendant outcomes up to 2016, we also limit our sample to cases processed 2005–2011, with a sample size of 42,183. Recidivism is measured as an indicator for having one or more charges registered by the police against the defendant for new crime(s). Almost 70% of the defendants in our sample are re-charged again in the five-year period post-decision, implying a fairly high recidivism rate. Future employment is measured by an indicator for ever having annual earnings above 1 SGA threshold in the Norwegian social security system in the five-year period post-decision. Only 56% of the defendants in our sample are ever employment. Similarly, we calculate each defendant's total earnings over the sample period.

²¹We use the fact that if $x \sim \mathcal{N}(0, \sigma)$ then $E[x|x > 0] = \frac{\sigma}{\sqrt{2\pi}}$.

Table 2.9: *The Effects of Judge Appeal Tendency on Defendant's Recidivism and Future Employment.*

	Pr(Defendant Ever Charged Within Five Years After Decision)	Pr(Defendant Ever Employed Within Five Years After Decision)	Defendant's Total Earnings Within Five Years After Decision (in 1000 NOK)
	(1)	(2)	(3)
A. Judge Appeal Tendency & Incarceration Stringency			
Appeal Tendency	0.001 (0.045)	-0.018 (0.058)	33.94 (92.11)
Incarceration Stringency	-0.131*** (0.043)	-0.030 (0.043)	6.26 (77.11)
Controls for X'_{ik} and α_{ct}	✓	✓	✓
B. Judge Appeal Tendency & Incarceration Stringency – Controlling for Other Judge Characteristics			
Appeal Tendency	-0.005 (0.046)	0.028 (0.054)	89.89 (90.93)
Incarceration Stringency	-0.116** (0.055)	-0.040 (0.063)	34.34 (104.6)
Controls for X'_{ik} and α_{ct}	✓	✓	✓
Controls for Judge Stringency	✓	✓	✓
Controls for Judge Disagreement	✓	✓	✓
Dependent Mean	0.70	0.56	555.7
Number of Cases	42,183	42,183	42,183

Note: Sample consists of non-confession criminal cases processed 2005-2014. All estimations include controls for court x court entry year FEs and all defendants and case characteristics listed in Table 2.1. Panel B also add controls for assigned judge's leave-out disagreement tendency, leave-out case processing time, judicial tenure and past caseload. Standard errors are two-way clustered at judge and defendant level. *p<0.1, **p<0.05, ***p<0.01.

We estimate our regression model on each of the three defendant outcomes. The results

from this exercise are provided in Table 2.9. This analysis reveals that the leave-out appeal tendency of the assigned judge neither affects defendants' recidivism or their future labor market outcomes. In contrast, we find that defendants assigned to judges with higher incarceration stringency have lower recidivism, consistent with the findings reported by Bhuller *et al.* (forthcoming). Based on this analysis, we thus do not find that high appeal tendency judges positively or adversely affect defendants' outcomes.

2.6 Conclusion

In this paper we have exploited the random assignment of cases to judges to show that there exists large and systematic differences in the appeal tendencies across judges. Furthermore, we have shown that differences in the rate at which judges convict, incarcerate, or impose long sentences does not explain all these differences. Finally, we found that judges with lower average case processing time, judges with preferences estimated to deviate from their peers, and judge who has spent a long time in the same court tend to have higher appeal rates. We see several directions to take this line of research further. First, a natural question to ask is which types of deviations in judicial preferences tend to be associated with higher appeal rates? Second, while our evidence imply that appeals must be at least partly explained by judicial errors or deviant judge preferences, we did not provide estimates for what share of all appeals can be attributed to this reason, as opposed to other reasons such as the design of the appeals system or litigant characteristics that might cause appeals even when trial court judges do not make mistakes or have deviant preferences. It should be possible to at least put some bounds on the share of appeals attributed to each channel. Finally, we believe we can get a better sense of the extent to which lower appeal tendencies matter for judicial appointments by matching lists of applicants for judicial positions to data on appeal tendencies.

Chapter 3

Coordinated Enforcement¹

3.1 Introduction

In many settings, such as those considered in principal agent problems (Laffont and Martimort, 2009) or classic models of crime and deterrence (Becker, 1968), it is presumed courts, judges, and the police enforce contracts and laws as dictated; the incentives of the enforcer typically can be safely ignored, or at least considered separately from the question of optimal deterrence. However, there are many instances in which enforcers only find it in their interest to punish transgressions when expected to (e.g., out of fear of being punished themselves if they do not punish) or when they expect others to (because it is advantageous to be on the ‘winning’ side of history). Some examples include revolutionaries overthrowing abusive regimes, protesters defending civil rights and liberties, branches of government checking each others’ power, companies colluding to restrict output, an on-going relationship between employer and an employee (e.g., Baker *et al.*, 2002; Levin, 2003), and nations enforcing international treaties and laws. Another prominent example is communities enforcing social norms, for instance to govern common resources (Ostrom, 2015; Acheson, 2003) resolve disputes (Ellickson, 1994), or maintain order in pre-state (Henrich, 2015; Boyd, 2017) or extra-legal societies such as the mafia, pirates (Leeson, 2007),

¹Co-authored with Moshe Hoffman, N. Aygun Dalkiran, and Erez Yoeli

and Maghribi traders (Greif, 1993). In fact, many of our moral intuitions arguably developed in, and, hence, have features shaped by such coordinated enforcement contexts (Pinker, 1997; Nowak and Sigmund, 2005; DeScioli and Kurzban, 2009; Henrich, 2015; Boyd, 2017).

We present a stylized model of such “coordinated enforcement” as follows. One player (the “agent”) takes an action that has direct payoff consequences to herself, and also influences the state of the world. Two additional players (or “enforcers”) receive a signal informative of the state of the world, then play a simultaneous coordination game, with payoffs to themselves that are independent of the state, their signals, and the agent’s action. The enforcers’ actions do, however, have payoff consequences for the agent, and, hence, can act as a sanction. We ask, for which subset of signals is it an equilibrium for those signals to trigger sanctioning, and, consequently, which actions can the agent be deterred from taking. While this model—and, in particular our use of a state-independent coordination game—is highly stylized, it still allows us to capture the key feature of coordinated enforcement—that enforcers are primarily motivated by others’ expectations—and is more tractable than, say, repeated games with private information. We provide the intuition for why our results would apply to repeated games, including dyadic ones, in the conclusion.

This model is intended to elucidate and formalize the essential influence of higher order beliefs on such settings. Intuitively, if enforcers are primarily motivated by others’ expectations, then what matters is not what their signal tells them about the state, but, rather, what their signal tells them about others’ signals. We formalize this intuition using the tools developed by Monderer and Samet (1989). In particular, the signals that trigger sanctioning need to be “dually $\bar{p}$ -evident”, meaning that, exactly when an enforcer gets a signal that triggers sanctioning, she expects the other to get such a signal with probability at least $\bar{p}$, where $\bar{p}$ corresponds to the risk-dominance of the coordination game. Hence, for the agent to be deterred, there needs to be a set of signals such that: this dually $\bar{p}$ -evident property holds (technically, defined according to the distribution that would be induced if the agent is properly deterred), and, these signals need to be sufficiently more likely to occur if the tempting but undesirable action is taken.

We also present two special cases that illustrate non-intuitive implications of the role of higher order beliefs. In the first, which is adapted from Hoffman *et al.* (2018), we employ a variation of the Global Games state-signal structure (Morris and Shin, 2001) to explain why, in coordinated enforcement settings, transgressions are typically defined based on categorical distinctions. For example, consider the international norm against chemical weapons. A violation of this norm is determined by the type of weapon used, and not, for example, by the number of needless deaths or the amount of suffering. We explain this phenomenon as follows. A categorical norm can be an equilibrium so long as, whenever one enforcer gets a signal that the categorical norm has been violated (for instance, she detects residue from a chemical weapons attack), her posterior that the other enforcer got a signal that the norm was violated sufficiently increases. Thereby, the dually $\bar{p}$ -evident property described above is satisfied. However, a “threshold norm” that prescribes sanctioning only for signals greater than some threshold (for instance, sanctioning whenever estimated casualties exceed, say, 10,000), is harder to support as an equilibrium, because one’s estimate that the other enforcer received a signal above the threshold may not sufficiently increase when one’s own signal moves from just below the threshold to just above it. Thereby, the requisite dually $\bar{p}$ -evident property would not be satisfied.

In our second special case, there are two states, 0 and 1, where we interpret state 1 as indicative of the behavior one is trying to deter. When the state is 1, each player receives the signal 1 with probability $1 - \epsilon$ and correlation of ρ . When the state is 0, players receive signal 0 with certainty. This simple state-signal structure allows us to formalize the effect of the state’s “observability” ($1 - \epsilon$) or of type II error (ϵ), and the degree to which information is “shared” or “public” (ρ). Crucially, conditional on receiving the signal, an enforcer is certain the state is 1 (her first order beliefs), yet, she is only confident the other player also received signal 1 (her second order beliefs) if either $1 - \epsilon$ or ρ is sufficiently high. Therefore, these parameters influence whether signal 1 can trigger sanctions even though they do not influence first order beliefs in the case that the enforcer receives signal 1. In contrast, in a standard deterrence setting, only first order beliefs matter.

We believe this result may elucidate the seemingly undue influence of plausible deniability—when a transgression goes unpunished or is seen as less bad simply because there is a plausibly innocent explanation, even if, in this instance, one knows that explanation is unlikely or does not apply. In such instances, one might know the innocent explanation does not apply but not be sufficiently confident others also know this, which, in our model, corresponds to the case where one gets signal 1, but $1 - \epsilon$ and ρ are both low. Plausible deniability arguably underlies many well-documented phenomena, such as: the omission-commission distinction (being more willing to allow harm to befall others than to take actions that cause comparable harm; Spranca *et al.* 1991; DeScioli *et al.* 2012); strategic ignorance (avoiding information about the negative social consequences of one's actions; (Dana *et al.*, 2007)); avoiding the ask (those who would otherwise help go out of their way to avoid being in a position to help; Dana *et al.* 2006; Andreoni *et al.* 2017); indirect harm (being judged less harshly if e.g., acting through an intermediary or subordinate; Greene *et al.* 2009); the means-byproduct distinction (being more willing to allow harm to befall others as an anticipated byproduct of one's actions than as a necessary means to an end; Greene *et al.* 2009; DeScioli *et al.* 2012); and “small” cheats (being more willing to lie or cheat if an innocent explanation exists, e.g., stealing a pen versus a dollar from the office; Ariely *et al.* 2009; Ariely 2012). For each of these phenomena, the puzzling case occurs when enforcers know the omission, byproduct, indirect harm, *etc.* was intentionally and knowingly caused, yet treat these as distinct from commissions, byproducts, direct harm, *etc.*, presumably because even when one knows that the omissions, *etc.* was intentionally, knowingly caused, one is not sufficiently confident such intentionality, foreknowledge, *etc.* are known by others.

The paper proceeds as follows. We begin by relating our work to prior literatures. Then, we present our model. First, we focus on the enforcers' decision by omitting the agent's decision and presuming the probability over states is exogenously determined (Sec. 3.3.1). We use this setting to present a straightforward, but useful result characterizing which signals can trigger sanctioning. Next, we connect this result with higher order beliefs

(Sec. 3.3.2). Then, we add the agent’s choice (Sec. 3.3.3), allowing us to formalize the effect of coordinated enforcement on deterrence, which yields our main theoretical result. Then, we develop our two special cases (Sec. 3.4.1 and 3.4.2). We conclude with discussions of the features of coordinated enforcement that are highlighted by these special cases, related literatures, the efficiency losses and perverse effects introduced by coordinated enforcement, as well as open questions (Sec. 3.5).

3.2 Relation to Prior Literatures

Our paper heavily builds off insights in the game theory literature on common knowledge and higher order beliefs (e.g., Aumann, 1976; Rubinstein, 1989; Monderer and Samet, 1989; Morris and Shin, 1997). Of closest relevance to this paper, Monderer and Samet (1989) formalized notions of higher order beliefs, developing the toolkit used herein, Rubinstein (1989) highlighted, in a particularly illuminating example (the Electronic Mail Game), the key role of common knowledge for coordination, and Carlsson and van Damme (1993); Morris and Shin (1998) developed a particular state-signal structure used to explain phenomena ranging from bank runs, currency attacks, and social unrest, and which we adapted in our categorical norms special case. Each of these spawned significant related literatures (e.g., Morris *et al.*, 1995; Morris and Shin, 2001; Frankel *et al.*, 2003; Morris, 2002, 2014).

Our work adds to the literature on higher order beliefs and coordination by characterizing the requirements on signals such that strategies can condition on signals in equilibrium, characterizing the role of higher order beliefs for coordination, spelling out the potential relevance for deterrence, and suggesting the potential for explaining puzzling social phenomena such as categorical norms (see also Hoffman *et al.*, 2018) and plausible deniability. With regards to the Global Games literature, our Sec. 3.4.1 differs as follows. Morris and Shin (1998), like us, presume a well ordered state-space and private signals of the state. However, Morris and Shin (1998) require payoffs to be state-dependent. Morris and Shin (1998)’s key question also differs from ours: Morris and Shin (1998) ask when is there a unique threshold equilibrium, whereas we ask when are there any threshold equilibria.

Notice that in our context, when no threshold equilibrium exists because the conditions of Thm. 3 are not satisfied, we expect strategies to not condition on signals, which would never be an equilibrium in Morris and Shin (1998), because they presume payoff dependence is sufficiently strong that each action is dominated in some region. Lastly, because the question we are asking is so different, the state-signal structures that satisfy the requirements of Thm. 3 do not, to the best of our knowledge, correspond to any important characterization in the Global Games literature.

Our paper also builds on less formal work that argues for the pertinence of higher order beliefs to interesting social behaviors. For instance, Chwe (2013) suggests that higher order beliefs explains symbolic displays, as well as the role of public events for sparking revolutions, Pinker *et al.* (2008) suggests that higher order beliefs can help explain indirect speech, and DeScioli and Kurzban (2009) suggests that higher order beliefs can help us understand quirks of morality like the omission-commission distinction. Our work adds to this work by adding formalism, such as our use of state-signal structures to formalize private information, triggering strategies to formalize what it means to be “able to coordinate”, and observability and correlation to model plausible deniability.

3.3 A Model of Coordinated Enforcement

3.3.1 Characterizing Which Signals Can Trigger Sanctions

There are two players. A state of the world, ω is randomly chosen from the set Ω according to the prior probability μ . In Sec. 3.3.3, we will consider what happens when the state is influenced by the actions of a third player, which we call the agent. Each player, i receives a private signal s_i drawn from the set S , according to the joint distribution $f_\omega(s_1, s_2)$, which is conditional on the state. Together, $(\Omega, \mu, S, \{f_\omega\}_{\omega \in \Omega})$ comprise the state-signal structure.

After a state, ω , is drawn according to μ , and signals, s_1, s_2 ,² are drawn according to

²To aid readability, we will treat s_i as an arbitrary signal realization and not as a random variable, as we do for ω .

f_ω , players play a simultaneous coordination game, G (Fig. 3.1), where playing A will be interpreted as sanctioning once we introduce the agent in Sec. 3.3.3. Players can condition on their signal. Note that the payoffs to sanctioning are presumed to not depend on the state or signals. Let $\bar{p} = \frac{d-b}{(d-b)+(a-c)}$ be the risk dominance of G . Its interpretation: so long as player $-i$ is playing A with probability greater than or equal to $\bar{p}$, player i prefers to play A . Together, G and $(\Omega, \mu, S, \{f_\omega\}_{\omega \in \Omega})$ induce a *coordination game with signals*, Γ .

	A	B
A	(a, a)	(b, c)
B	(c, b)	(d, d)

where $a > c$ and $d > b$

Figure 3.1: The Coordination Game, G

Let $\sigma_i : S \rightarrow \{A, B\}$ represent player i 's strategy. We interpret s as *triggering sanctions* for player i if $\sigma_i(s) = A$. We restrict our analysis to pure strategies. We assume player i maximizes expected payoffs for every signal s_i , given μ, f_ω , and σ_{-i} , and define a Bayesian Nash Equilibrium of Γ as follows:

Definition 1. A strategy profile $\sigma = (\sigma_1^*, \sigma_2^*)$ is a Bayesian Nash Equilibrium (B.N.E.) of Γ iff for $i \in \{1, 2\}$, for all $s_i \in S$, $EU_i(\sigma_i^*(s_i), \sigma_{-i}^*|s_i) \geq EU_i(a_i, \sigma_{-i}^*|s_i)$ for all $a_i \in \{A, B\}$,

$$\text{where } EU_i(\sigma_i^*(s_i), \sigma_{-i}^*|s_i) \equiv \int_{\omega \in \Omega} \int_{s' \in S} u_i(\sigma_i(s_i), \sigma_{-i}(s')) df_\omega d\mu$$

and $u_i(\cdot, \cdot)$ is given by G .

We refer to a subset of $\Omega \times S \times S$ as an *event*. Sometimes, in an abuse of notation, we specify such events by a signal or set of signals for a player, i . For example, the event that player i gets signal s'_i is $\{(\omega, s_1, s_2) \in \Omega \times S \times S | s_i = s'_i\}$. We refer to this event as $\mathbf{s}'_i$. Similarly, the event that player $-i$ gets a signal in S_{-i}^* is $\{(\omega, s_1, s_2) \in \Omega \times S \times S | s_{-i} \in S_{-i}^*\}$. We refer to this event as $\mathbf{S}_{-i}^*$. Thus, for two events, $\mathbf{E}$ and $\mathbf{F}$, we can calculate:

$$P(\mathbf{E}|\mathbf{F}) = \frac{\int_{(\omega, s_1, s_2) \in \mathbf{E} \cap \mathbf{F}} df_\omega d\mu}{\int_{(\omega, s_1, s_2) \in \mathbf{F}} df_\omega d\mu}$$

We then define p -beliefs, as in Monderer and Samet (1989):

Definition 2. For an event $\mathbf{E} \subseteq \Omega \times S \times S$, we say that player i p -believes $\mathbf{E}$ at $\mathbf{s}'_i$ given μ iff:

$$P(\mathbf{E}|\mathbf{s}'_i) \geq p$$

Our main result is that, for any subsets of signals, S_1^*, S_2^* , the strategy where player i plays A whenever she gets a signal in S_i^* is a B.N.E. if and only if: whenever a player receives a signal in S_i^* , she $\bar{p}$ -believes that the other received some signal in S_{-i}^* , and whenever she receives a signal not in S_i^* , she $(1 - \bar{p})$ -believes that the other did not receive a signal in S_{-i}^* . Formally:

Theorem 1. Consider the coordination game with signals $\Gamma = (G, \Omega, S, \mu, F)$. Consider any $S_1^*, S_2^* \subseteq S$ and define the strategy $\sigma_i^*(s_i) = A$ iff $s_i \in S_i^*$. The strategy profile (σ_1^*, σ_2^*) is a B.N.E. of Γ iff for $i = 1, 2$:

$$s_i \in S_i^* \Rightarrow \text{player } i \text{ } \bar{p}\text{-believes } \mathbf{S}_{-i}^* \quad (3.1)$$

$$s_i \notin S_i^* \Rightarrow \text{player } i \text{ } (1 - \bar{p})\text{-believes } \neg \mathbf{S}_{-i}^* \quad (3.2)$$

Proof. All proofs are presented in Appendix C. □

3.3.2 The Connection With Higher Order Beliefs

In this section, we relate Thm. 1 to higher order beliefs. As with Def. 2, we will continue to adapt notation and insights from Monderer and Samet (1989) to our setting. We begin by defining the set of all state-signal combinations in which an event $\mathbf{E} \subseteq \Omega \times S \times S$ is p -believed by player i :

Definition 3. For $i \in \{1, 2\}$ and $\mathbf{E} \subseteq \Omega \times S \times S$,

$$B_i^p(\mathbf{E}) \equiv \{(\omega, s_1, s_2) \in \Omega \times S \times S \mid P(\mathbf{E}|\mathbf{s}_i) \geq p\}$$

If everyone p -believes an event whenever it occurs, we call it p -evident:

Definition 4. *The event $\mathbf{E}$ is p -evident if, for $i \in \{1, 2\}$, $\mathbf{E} \subseteq B_i^p(\mathbf{E})$*

We also refer to the signals (S_1, S_2) as p -evident if the event $\Omega \times S_1 \times S_2$ is p -evident.

The following corollary relates Thm. 1 to $\bar{p}$ -evident events:

Corollary 1. *Consider the coordination game with signals $\Gamma = (G, \Omega, S, \mu, F)$. Consider any $S_1^*, S_2^* \subseteq S$ and define the strategy $\sigma_i^*(s_i) = A$ iff $s_i \in S_i^*$. Let $\mathbf{E}^* = \Omega \times S_1^* \times S_2^*$ and $\mathbf{F}^* = \Omega \times \neg S_1^* \times \neg S_2^*$. Then, the strategy profile (σ_1^*, σ_2^*) is a B.N.E. of Γ iff $\mathbf{E}^*$ is $\bar{p}$ -evident and $\mathbf{F}^*$ is $(1 - \bar{p})$ -evident.*

To relate this result to our claim in the introduction that the signals that trigger sanctioning need to be “dually $\bar{p}$ -evident”, we can define (S_1^*, S_2^*) as dually $\bar{p}$ -evident if $\mathbf{E}^* = \Omega \times S_1^* \times S_2^*$ and $\mathbf{F}^* = \Omega \times \neg S_1^* \times \neg S_2^*$ are $\bar{p}$ -evident and $(1 - \bar{p})$ -evident, respectively.

3.3.3 The Connection With Deterrence

Now, we wish to relate Thm. 1 to deterrence. Recall, we are motivated by the question: when deterrence occurs via coordinated enforcement, what actions can be deterred, and what signals must trigger sanctions? There are now three players: an agent, and two enforcers. The agent chooses an action, which influences the state of the world as well as her payoffs. The enforcers observe signals of the state as in Sec. 3.3.1, then choose whether to sanction the agent. However, unique to our framework, while enforcers’ choices have repercussions for the agent, their own payoffs are only affected by the desire to coordinate with each other, which we represent by the coordination game G .

Formally, the agent chooses an action $a \in \mathbf{A}$. Each action induces a distribution, μ_a , on the states of the world, Ω . The state of the world also yields payoffs to the agent, $v : \Omega \rightarrow \mathbb{R}$. As in Sec. 3.3.1: for a given $\omega \in \Omega$, each enforcer, $i = 1, 2$, gets signal $s_i \in S$ according to the joint distribution $f_\omega(s_1, s_2)$; and, each enforcer chooses an action from $\{A, B\}$ as a function of her signal, and gets payoffs according to the coordination game G . Player i ’s strategy is

again represented by σ_i . Here, we interpret choosing A as sanctioning the agent. The agent incurs cost $c(\cdot)$ from being sanctioned, which we will define in the next paragraph.

Let $n : \{A, B\} \times \{A, B\} \rightarrow \{0, 1, 2\}$ be the count of A s. E.g., $n(A, B) = 1$, and $n(B, B) = 0$. We define the cost function $c(\cdot)$ as $c : \{0, 1, 2\} \rightarrow \mathbb{R}$, and assume it is monotonically increasing. Therefore, when signals s_1 and s_2 are realized and strategies σ_1 and σ_2 are utilized, the cost from being sanctioned is $c(n(\sigma_1(s_1), \sigma_2(s_2)))$. In any state $\omega \in \Omega$, the expected cost from being sanctioned is $\int_{(s_1, s_2) \in S \times S} c(n(\sigma_1(s_1), \sigma_2(s_2))) df_\omega$. Thus, for any distribution μ_a , the expected costs from sanctioning are $\int_{\omega \in \Omega} \int_{(s_1, s_2) \in S \times S} c(n(\sigma_1(s_1), \sigma_2(s_2))) df_\omega d\mu_a$.

Finally, we assume the agent chooses a to maximize her expected payoffs net of sanctions, given σ_1 and σ_2 :

$$V(a, \sigma_1, \sigma_2) = \int_{\omega \in \Omega} \left(v(\omega) - \int_{(s_1, s_2) \in S \times S} c(n(\sigma_1(s_1), \sigma_2(s_2))) df_\omega \right) d\mu_a$$

Definition 5. We say action $a^* \in A$ can be enforced if there exists a σ_1^*, σ_2^* such that $(a^*, \sigma_1^*, \sigma_2^*)$ is a perfect B.N.E..

Note that by perfect B.N.E., we mean the natural extension of B.N.E. presented in Def. 1, in which the agent best responds given what she expect enforcers to do, and each enforcer best responds given the expected behavior of the agent and the other enforcer's behavior, where expectations are all consistent with Bayes' Rule wherever possible.

Theorem 2. Action a^* can be enforced if and only if there exists $S_1^*, S_2^* \subseteq S$ such that:

$$\text{for } i = 1, 2 : s_i \in S_i^* \Leftrightarrow i \text{ } \bar{p}\text{-believes } \mathbf{S}_{-i}^* \text{ given } \mu_{a^*} \quad (3.3)$$

$$a^* \in \arg \max_{a \in A} V(a, \sigma_1^*, \sigma_2^*), \text{ where for } i \in \{1, 2\}, \sigma_i^*(s_i) = A \text{ iff } s_i \in S_i^* \quad (3.4)$$

Notice condition 3.4 is the same deterrence condition that would exist in a framework where enforcers could commit to acting in advance; this condition requires that any action with higher direct payoffs to the agent than a^* makes the signals that induce sanctioning sufficiently more likely so as to offset the gains in direct payoffs. Condition 3.3 is the same

as in Sec. 3.3.1, except that, now, the probability measure μ is not exogenously given, but the one induced by action a^* . Notice that condition 3.3 is unique to this framework that involves coordination. Thm. 2 highlights a trade-off unique to deterrence settings that involve coordination: conditioning sanctions on a signal that is indicative that the agent chose an undesired action can help to deter that action, but might require sanctioning on other signals which would, otherwise, not be needed.

3.4 Special Cases

3.4.1 Categorical Norms

An Example With Uniform Distributions of States and Signals

We now describe a simple state-signal structure, reminiscent of the state-signal structure in the Global Games literature (Morris and Shin, 2001), that will allow us to highlight the role of coordinated punishment in preventing “threshold norms”. The setup and results in Sec. 3.4.1 are adapted from Hoffman *et al.* (2018).

Consider the following coordination game with signals $\Gamma = (G, \Omega, S, \mu, F)$: $\Omega = \mathbb{R}$; μ represents the uniform distribution with improper priors; $S = \mathbb{R}$, and $f_\omega(s_1, s_2) = \frac{1}{4\epsilon^2}$ for $s_1, s_2 \in [\omega - \epsilon, \omega + \epsilon]$ and 0 otherwise; and G is the same coordination game as in Sec. 3.3.1.

We begin by defining a threshold norm.

Definition 6. Consider an $\underline{s} \in S$. Define the strategy $\sigma_i^{\underline{s}}(s_i) = A$ for $s_i > \underline{s}$, and B for $s_i < \underline{s}$. We say $(\sigma_1^{\underline{s}}, \sigma_2^{\underline{s}})$ is a threshold norm at $\underline{s}$.

Now, we show that, unless the risk-dominance of the coordination game, $\bar{p}$, is non-generically equal to .5, any threshold norm is unstable.

Proposition 1. For any $\underline{s} \in S$, the threshold norm at $\underline{s}$ is a B.N.E. iff $\bar{p} = .5$.

The intuition for this result is as follows. For the threshold norm at $\underline{s}$ to be upheld, a player must find it in her interest to impose sanctions if and only if her signal is greater than $\underline{s}$. However, there will be signals quite close to $\underline{s}$ at which she will want to deviate, since at

such signals, she believes the other player is roughly 50% likely to sanction, which will not, generally, be the point at which she is indifferent.

Categorical Norms: A Result for Arbitrary Distributions of States and Signals

We can generalize the setup in Sec. 3.4.1 to get a better sense for which state and signal distributions, and at which threshold values, it will be possible for threshold norms to be B.N.E.s. Consider the following coordination game with signals $\Gamma = (G, \Omega, S, \mu, F)$: $\Omega = \mathbb{R}$; μ represents an arbitrary probability distribution on Ω ; $S = \mathbb{R}$, and the signal distributions, $f_\omega(s_1, s_2)$ satisfy the Monotone Likelihood Ratio Property (MLRP), i.e. for any $\omega, \omega' \in \Omega$ and any $s, s' \in S$, for $i = 1, 2$, if $\omega < \omega'$ and $s < s'$, $\frac{f_\omega(s)}{f_\omega(s')} < \frac{f_{\omega'}(s)}{f_{\omega'}(s')}$; and G is the same coordination game as in Sec. 3.3.1. The definition of a threshold norm carries over from Sec. 3.4.1 to this setup unchanged, though, for convenience, we now assume that the threshold $\underline{s}$ does not occur at an atom.

Theorem 3. *For any $\underline{s} \in S$, a threshold norm at $\underline{s}$ is a B.N.E. for $i \in \{1, 2\}$ iff:*

$$\sup_{s_i < \underline{s}} P(s_{-i} < \underline{s} | s_i) \geq 1 - \bar{p} \geq \inf_{s_i > \underline{s}} P(s_{-i} < \underline{s} | s_i)$$

Thm. 3 suggests three mechanisms by which threshold norms are supported in equilibrium. The first two apply to continuous distributions of states and signals. For such pairs of distributions, for any threshold $\bar{s}$, the *R.H.S.* and *L.H.S.* of the condition in Thm. 3 must be equal. Our first mechanism applies when pairs of distributions such that the *R.H.S.* and *L.H.S.* are constant in $\bar{s}$. In this case, only a particular $\bar{p}$ can satisfy the theorem's inequalities, regardless of $\bar{s}$. For instance, as we saw in Sec. 3.4.1, for uniformly distributed states and signals that are symmetrically distributed about the state, the *R.H.S.* and *L.H.S.* are equal to .5 for any $\bar{s}$, so this is the only value of $\bar{p}$ that satisfies Thm. 3's inequalities. Thus, this mechanism permits threshold equilibria only for non-generic $\bar{p}$, but for this $\bar{p}$, it is possible to choose any threshold, $\bar{s}$.

Turning to our second mechanism, for some pairs of continuous distributions, the *R.H.S.* and *L.H.S.* of Thm. 3 vary with $\bar{s}$. So long as they do so monotonically, the *R.H.S.* and *L.H.S.* will equal $\bar{p}$ at at most one $\bar{s}$. An example that fits this case is when states are normally distributed and noise is symmetrically distributed about, and independent of, the state. This mechanism, unlike the previous one, may allow for threshold equilibria for generic values of $\bar{p}$. However this mechanism does not allow for generic values of $\bar{s}$; that is, while threshold equilibria may exist for a wide range of coordination games, for each game, the threshold, $\bar{s}$, will be restricted to a particular value—one that may not, for instance, optimize deterrence.

A third means of obtaining a threshold equilibrium is for the distribution of states to have atoms, since atoms allow for a ‘gap’ between the *R.H.S.* and the *L.H.S.* of the condition in Thm. 3, into which $\bar{p}$ can fit. The range of $\bar{p}$ s will depend on the mass on the atoms: the greater the mass, the greater the range, and thus of coordination games for which threshold equilibria exist. In addition to permitting a range of $\bar{p}$ s, this mechanism can permit one to choose from multiple $\bar{s}$ s: namely, if there are multiple atoms, then one can choose at which one to set $\bar{s}$. This analysis implies that distributions that are ‘more discrete’ (i.e. those that have fewer values, each with a larger mass) ‘enable’ threshold norms. And, even when distributions are not exactly continuous, but are simply ‘close’ to continuous (i.e. there are signals with positive, but arbitrarily small mass) threshold norms will only be possible for coordination games with risk dominance values within an arbitrarily small range.

For more details on these points, see Hoffman *et al.* (2018).

3.4.2 The Role of Observability and Correlation

We now describe a simple state-signal structure which will allow us to convey the role of “observability” and “correlation” of signals. Consider the following coordination game with signals, $\Gamma = (G, \Omega, S, \mu, F)$. The possible states of the world are $\Omega = \{0, 1\}$, with prior $\mu(1) = \mu_1$ for some fixed $\mu_1 \in (0, 1)$. The set of possible signals is $S = \{0, 1\}$, and the signal distribution is as follows:

- $P(s_i = 0 | \omega = 0) = 1$, i.e. no false positives.

- $P(s_i = 1|\omega = 1) = 1 - \epsilon$, which can be interpreted as the ‘observability’ of state 1, and ϵ is the frequency of false negatives.
- $P(s_i = 1|s_{-i} = 1, \omega = 1) = r$, which can be interpreted as a measure of how ‘shared’ or ‘public’ are players’ observations of state 1. Note, when players’ signals are independent, conditional on $\omega = 1$, then $P(s_i = 1|s_{-i} = 1, \omega = 1) = P(s_i = 1|\omega = 1)$ and, hence, $r = 1 - \epsilon$. Whereas when they are correlated, conditional on $\omega = 1$, then $r > 1 - \epsilon$. This can be converted to the standard measure of conditional correlation, ρ , between s_i and s_{-i} given $\omega = 1$, using the following formula $\rho = \frac{r(1-\epsilon)-(1-\epsilon)^2}{1-\epsilon-(1-\epsilon)^2}$.³

G is the same coordination game as in Sec. 3.3.1.

Proposition 2. *Let $S_1^* = S_2^* = \{1\}$ and define the strategy $\sigma_i^*(s_i) = A$ iff $s_i \in S_i^*$. The strategy profile (σ_1^*, σ_2^*) is a B.N.E. of Γ iff:*

$$1 - \epsilon(1 - \rho) \geq \bar{p} \quad (3.5)$$

$$\frac{\mu_1 \epsilon (1 - \epsilon) (1 - \rho)}{\mu_1 \epsilon + (1 - \mu_1)} \leq \bar{p} \quad (3.6)$$

Condition 3.5 and 3.6 come from conditions 3.1 and 3.2, respectively. The interpretation of condition 3.5 is as follows: it is possible to condition behavior on one’s signal only if signals are not too noisy or signals are highly correlated. Note that condition 3.5 is equivalent to $r \geq \bar{p}$. However, the above formulation breaks the informativeness of player i ’s signal on player $-i$ ’s signal into two pieces: the part that comes from players’ signals being correlated, conditional on the state, and the part that comes from player i ’s signal being indicative of the state, which directly effects the distribution of player $-i$ ’s signal.

³The standard definition of correlation is: $\rho_{XY} = \frac{E((X-\mu_X)(Y-\mu_Y))}{\sigma_X \sigma_Y}$. In our context: $X = (s_i|\omega = 1)$, $Y = (s_{-i}|\omega = 1)$, $\mu_{s_i|\omega=1} = \mu_{s_{-i}|\omega=1} = 1 - \epsilon$, and $\sigma_{s_i|\omega=1} = \sigma_{s_{-i}|\omega=1} = \sqrt{(1 - \epsilon) - (1 - \epsilon)^2}$. This yields: $\rho_{s_i|\omega=1, s_{-i}|\omega=1} = \frac{r(1-\epsilon)-(1-\epsilon)^2}{1-\epsilon-(1-\epsilon)^2}$.

Condition 3.6 will not bind if ϵ or μ_1 are sufficiently small. This condition is less likely to bind because we assumed there are no false positives.

We note that when $s_i = 1$, player i knows with certainty that $\omega = 1$, but when ϵ is high and ρ is low, she will not be sufficiently confident that the other player got the same signal. Hence, we expect players to be less able to condition on such signals, compared to signals where ϵ is lower or ρ is higher. We use this to explain the seemingly undue effect of plausible deniability in settings with coordinated enforcement. For instance, in the omission-commission distinction, the puzzling case is when enforcers know an omission is intended, but still do not sanction as much as if it were a commission (Spranca *et al.*, 1991). We would expect this to be the case if signals indicative of intentions have higher rates of false negatives or are less public than signals indicative of actions.

3.5 Discussion

We considered the setting of coordinated enforcement, which we argued is a distinct setting for deterrence than that traditionally considered in the literature. We presented a stylized model of such coordinated enforcement, then characterized which signals can trigger sanctions and what actions can be deterred, as well as the crucial role of higher order beliefs. We also presented two special cases to illustrate some unique and puzzling features of coordinated enforcement settings. In these, we showed that: (1) it is difficult to enforce norms that condition on continuous variables, and, (2) it is difficult to enforce norms that condition on signals which are not sufficiently observable or shared.

In this section, we discuss implications of our framework. The appendix includes additional discussion of practical prescriptions, and of implications for models of crime and punishment, and for characterizing pro-social preferences. Before we begin, we emphasize that the implications of coordinated enforcement might apply not just to our conscious deliberation but also to our moral intuitions and social preferences, since these are shaped by learning and evolutionary processes, and thus are expected to show features consistent with Nash equilibria (Weibull, 1997). Of course, such intuitions and preferences may, to

some extent, spill over into settings that do not involve coordinated enforcement.

3.5.1 Comparative Statics

Our work suggests the following comparative static: by identifying domains in which coordinated enforcement is less important, one can identify domains in which phenomena such as categorical norms or plausible deniability are expected to be less pertinent. For instance, we expect such phenomena to show up less in domestic than in international law, since the former can be unilaterally enforced by, e.g., a law enforcement agency, whereas the latter must be coordinated amongst multiple countries. In the domain of altruism, under our usual presumption that social preferences are shaped by coordinated enforcement, we expect such phenomena to show up less amongst family than amongst strangers, since altruism towards family is more driven by kin selection, whereas towards strangers, it is primarily driven by norm enforcement or indirect reciprocity (Nowak and Sigmund, 2005; Nowak, 2006; Boyd, 2017). Indeed, we find that people are more sensitive to efficacy and attend less to the omission-commission distinction amongst kin than amongst strangers (Burum *et al.*, 2018). And, we expect such phenomena to show up less for private decisions where one's optimal action is primarily dictated by first order beliefs, like who to work with, date, or avoid. For example, people take age, height, or attractiveness into account in such cases, even though these variables are continuous. And, people avoid those who are strategically ignorant or who transgress by omission, even if they do not punish them.

3.5.2 Welfare Implications

Another implication of our model is that first-best outcomes cannot always be enforced. This is illustrated by both our special cases. Often, it is socially optimal to condition on continuous variables. For instance, one might wish to have a norm that prescribes only contributing to the public good (say, by recycling) when it is efficient to do so (the environmental benefits outweigh the effort and cost). However, efficiency is a continuous variable, so, instead, our norms prescribe that we always recycle certain categories of items,

even when it, e.g., requires repeatedly, and wastefully, rinsing a container. Our second special case highlights that it can be difficult to enforce socially optimal behavior when it, e.g., requires penalizing inaction. For instance, it might be socially optimal, but difficult, to incentivize people to intervene to prevent a crime, or to check abuses of power by the executive branch.

Another counter-intuitive implication is that more information can hamper our ability to deter. Again, this can be seen in both our special cases. In the first special case, it is possible to deter using a coarse signal that meets the condition in Thm. 3, whereas refining that signal might prevent the condition in Thm. 3 from holding. In the second special case, enabling players to better, but idiosyncratically, observe the state, can hamper coordinated enforcement. We note for contrast that, in a standard deterrence framework, more information can only (weakly) facilitate deterrence.

3.5.3 Prescriptions

In this section, we discuss some prescriptions implied by our framework. First, when promoting pro-social behavior, one should remove plausible excuses, as Andreoni *et al.* (2017) did when they had Salvation Army volunteers stand in front of both the doors to the supermarket, instead of just one, eliminating donors' ability to go out the other door then, e.g., claim they just did not see the volunteers. Another practical prescription has to do with combating ineffective charitable giving. Our model suggests that in order for norms that encourage more effective giving to develop, one must make indications of higher efficacy both categorical and public, as the non-profit evaluator GiveWell does, by certifying highly effective charities. A third is that, to the extent that quirks like categorical norms or plausible deniability influence domestic law, this is likely due to "spillovers" from the coordinated enforcement context in which our moral intuitions develop to contexts in which enforcement does not require coordination, and, thus, is not benefited by such quirks. For instance, the law allows doctors to prescribe palliative care (to let someone die, an omission), but not to assist them in suicide (a commission). We suggest that one should either outlaw

both, or permit both, since the reason for the distinction is likely such a spillover. A fourth is that, to alleviate the inefficiencies caused by coordinated enforcement in settings like international law, one can create organizations with the power to unilaterally enforce norm violations—or, at least, to investigate them and publicly announce their findings—like the International Criminal Court at the Hague, or the International Atomic Energy Agency. Such an organization should be tasked with, for example, sanctioning or investigating rulers whose omissions enable genocide, or who causes excessive civilian casualties. A fifth relates again to international norms, as well as to checks and balances. Our model suggests that, to be sustainable, checks and balances and international norms should be conditioned on events that, whenever they occur, are evident. E.g., an international norm against chemical weapons is sustainable because it is categorical and there is common agreement on the definition of a chemical weapon, whereas a violation of the Emoluments Clause of the Constitution proved to be an ineffective check, perhaps because the definition of a violation had no precedent, and had not otherwise been commonly agreed upon.

3.5.4 Implications for Crime and Punishment

In this section, we note an implication of our framework for standard models of crime and punishment. These prescribe that punishment ought to increase proportionately to the gains from crime and inversely to the probability of getting caught (Becker, 1968). However, punishment, and our corresponding moral intuitions, are relatively insensitive to these variables (Carlsmith *et al.*, 2002). This has been used to argue that our judicial system and moral intuitions are not designed for optimal deterrence. However, we would argue that they may be second best, given the constraints imposed by coordinated enforcement. As a result, it is more difficult for us to deter crimes that are unlikely to be detected.

3.5.5 Characterizing Pro-social Preferences

Our model also has implications for characterizing pro-social preferences. Consider Becker (1974)'s canonical characterization of altruism, in which the welfare of other individuals, or

the amount of a public good provided, enter directly into individuals' utility. Our model suggests a different characterization, at least in contexts where altruistic preferences are shaped by coordinated enforcement: we expect people to get utility from contributing, particularly in ways that are observable and categorical, and when defecting would be hard to plausibly justify. Indeed, people are sensitive to whether they are being observed (Kraft-Todd *et al.*, 2015), get pleasure from the act of giving (Andreoni, 1990) which is categorical, are insensitive to multipliers (Karlan and List, 2006) and scope (Desvousges *et al.*, 1992; Kahneman *et al.*, 1999) which are continuous variables, and sensitive to the ability to "avoid the ask" (Dana *et al.*, 2006; DellaVigna *et al.*, 2012; Andreoni *et al.*, 2017) or remain "strategically ignorant" about the ramifications of their actions (Dana *et al.*, 2007). Consider, also, subsequent models such as Andreoni (1989); Fehr and Schmidt (1999); Charness and Rabin (2002), which mold the utility function to fit puzzling behavioral results, namely, by e.g., adding a warm glow from giving, or an aversion to inequity. Our model highlights a problem with this approach: the degree of warm glow is not expected to be a fixed parameter, but, rather, depend on the degree of observability and plausible deniability.

3.5.6 Intuition for Repeated Games and Future Works

This paper is meant to illustrate the value of the coordinated enforcement framework. It is natural to ask how our results generalize beyond the stylized coordination game, and especially to repeated games and relational contracts. Intuitively, in dyadic repeated interactions, higher order beliefs are expected to matter as well for the following reason. Consider an equilibrium in which transgressions are deterred through punishment. If punishing requires paying an immediate cost, the punisher will be tempted not to punish. Thus, in any perfect equilibrium, she must anticipate being penalized herself for not punishing. Such a penalty might, for example, occur if the other player is more likely to transgress whenever her original transgression goes unpunished. This is where higher order beliefs fit in. Whether the player wishes to punish a transgression depends on whether she anticipates that the other player also believes it was a transgression and will thus transgress

again if the original transgression goes unpunished. This logic iterates to higher orders as well (does the other player believe that she believes a transgression occurred, and will transgress if it goes unpunished? *Etc.*). To ensure this argument is not erroneous, and determine the conditions it depends on, this argument must be formalized. Indeed some preliminary work seems contradictory with the above intuition (Mailath and Morris, 2002, 2006).

Some additional further work we think would be valuable includes: fully characterizing the implications regarding welfare and information discussed above; considering more complicated extensive forms that, e.g., permit the agent to choose after receiving private information, and which might more directly map to phenomena such as strategic ignorance; and, considering more complicated information structures, e.g., to explore what counts as a public or private signal, perhaps by considering how information spreads on a network.

References

- ABRAMS, D. S., BERTRAND, M. and MULLAINATHAN, S. (2012). Do judges vary in their treatment of race? *The Journal of Legal Studies*, **41** (2), 347–383.
- ACHESON, J. M. (2003). *Capturing the commons: Devising institutions to manage the Maine lobster industry*. University Press of New England.
- AIZER, A. and DOYLE JR, J. J. (2015). Juvenile incarceration, human capital, and future crime: Evidence from randomly assigned judges. *The Quarterly Journal of Economics*, **130** (2), 759–803.
- AKHTARI, M., MOREIRA, D. and TRUCCO, L. (2018). Political turnover, bureaucratic turnover, and the quality of public services, unpublished.
- ALESINA, A. and FERRARA, E. L. (2014). A test of racial bias in capital sentencing. *American Economic Review*, **104** (11), 3397–3433.
- ANDREONI, J. (1989). Giving with impure altruism: Applications to charity and ricardian equivalence. *The Journal of Political Economy*, **97** (6), 1447–1458.
- (1990). Impure altruism and donations to public goods: a theory of warm-glow giving. *The Economic Journal*, **100** (401), 464–477.
- , RAO, J. M. and TRACHTMAN, H. (2017). Avoiding the ask: A field experiment on altruism, empathy, and charitable giving. *Journal of Political Economy*, **125** (3), 625–653.
- ARIELY, D. (2012). *The (honest) truth about dishonesty: How we lie to everyone—especially ourselves*. HarperCollins New York, NY.
- , BRACHA, A. and MEIER, S. (2009). Doing good or doing well? Image motivation and monetary incentives in behaving prosocially. *The American Economic Review*, **99** (1), 544–555.
- ARNOLD, D., DOBBIE, W. and YANG, C. S. (2018). Racial Bias in Bail Decisions. *The Quarterly Journal of Economics*, **133** (4), 1885–1932.
- AUMANN, R. J. (1976). Agreeing to disagree. *The Annals of Statistics*, pp. 1236–1239.
- AVIS, E., FERRAZ, C. and FINAN, F. (2018). Do Government Audits Reduce Corruption? Estimating the Impacts of Exposing Corrupt Politicians. *Journal of Political Economy*, **126** (5), 1912–1964.

- BAKER, G., GIBBONS, R. and MURPHY, K. J. (2002). Relational contracts and the theory of the firm. *The Quarterly Journal of Economics*, **117** (1), 39–84.
- BECKER, G. S. (1968). Crime and punishment: an economic approach. *The Journal of Political Economy*, **76** (2), 169–217.
- (1974). A theory of social interactions. *The Journal of Political Economy*, **82** (6), 1063–1093.
- BHULLER, M., DAHL, G. B., LØKEN, K. V. and MOGSTAD, M. (forthcoming). Incarceration, Recidivism, and Employment. *Journal of Political Economy*.
- BOHN, A. (2000). *Domsstolloven, Kommentarutgave [Law of Courts, Annotated Edition]*. Universitetsopplaget, Oslo (in Norwegian).
- BOYD, R. (2017). *A Different Kind of Animal: How Culture Transformed Our Species*. Princeton University Press.
- BURUM, B., HOFFMAN, M. and NOWAK, M. (2018). Hidden motives behind ineffective altruism. Harvard University Working Paper.
- CALONICO, S., CATTANEO, M. D. and TITIUNIK, R. (2014). Robust Nonparametric Confidence Intervals for Regression-Discontinuity Designs. *Econometrica*, **82** (6), 2295–2326.
- CARLSMITH, K. M., DARLEY, J. M. and ROBINSON, P. H. (2002). Why do we punish? deterrence and just deserts as motives for punishment. *Journal of personality and social psychology*, **83** (2), 284.
- CARLSSON, H. and VAN DAMME, E. (1993). Global games and equilibrium selection. *Econometrica*, **61** (5), 989–1018.
- CHARNESS, G. and RABIN, M. (2002). Understanding social preferences with simple tests. *Quarterly Journal of Economics*, **117** (3), 817–869.
- CHAVEZ, R. B. (2004). *The Rule of Law in Nascent Democracies: Judicial Politics in Argentina*. Stanford University Press.
- CHOI, S. J. and GULATI, G. M. (2004). Choosing the next supreme court justice: An empirical ranking of judge performance. *S. Cal. L. Rev.*, **78**, 23.
- CHWE, M. S.-Y. (2013). *Rational ritual: Culture, coordination, and common knowledge*. Princeton University Press.
- COLONNELLI, E., TESO, E. and PREM, M. (2017). Patronage in the Allocation of Public Sector Jobs, unpublished.
- DANA, J., CAIN, D. M. and DAWES, R. M. (2006). What you don't know won't hurt me: Costly (but quiet) exit in dictator games. *Organizational Behavior and Human Decision Processes*, **100** (2), 193–201.
- , WEBER, R. A. and KUANG, J. X. (2007). Exploiting moral wiggle room: experiments demonstrating an illusory preference for fairness. *Economic Theory*, **33** (1), 67–80.

- DELLAVIGNA, S., LIST, J. A. and MALMENDIER, U. (2012). Testing for altruism and social pressure in charitable giving. *The Quarterly Journal of Economics*, **127** (1), 1–56.
- DESCIOLI, P., ASAO, K. and KURZBAN, R. (2012). Omissions and byproducts across moral domains. *PloS One*, **7** (10), e46963.
- and KURZBAN, R. (2009). Mysteries of morality. *Cognition*, **112** (2), 281–299.
- DESVOUSGES, W. H., JOHNSON, F. R., DUNFORD, R. W., BOYLE, K. J., HUDSON, S. P. and WILSON, K. N. (1992). *Measuring nonuse damages using contingent valuation: An experimental evaluation of accuracy*, vol. 1992. Research Triangle Institute Research Triangle Park, NC.
- DI TELLA, R. and SCHARGRODSKY, E. (2013). Criminal recidivism after prison and electronic monitoring. *Journal of Political Economy*, **121** (1), 28–73.
- DIMITROVA-GRAJZL, V., GRAJZL, P., ZAJC, K. and SUSTERSIC, J. (2012). Judicial incentives and performance at lower courts: Evidence from slovenian judge-level data. *Review of Law & Economics*, **8** (1), 215–252.
- DOBBIE, W., GOLDIN, J. and YANG, C. S. (2018). The effects of pretrial detention on conviction, future crime, and employment: Evidence from randomly assigned judges. *American Economic Review*, **108** (2), 201–40.
- ELLICKSON, R. C. (1994). *Order without law*. Harvard University Press.
- FEHR, E. and SCHMIDT, K. M. (1999). A theory of fairness, competition, and cooperation. *The Quarterly Journal of Economics*, **114** (3), 817–868.
- FERRAZ, C. and FINAN, F. (2008). Exposing corrupt politicians: The effects of Brazil’s publicly released audits on electoral outcomes. *Quarterly Journal of Economics*, **123** (2), 703–745.
- and — (2011). Electoral Accountability and Corruption: Evidence from the Audits in Local Governments. *American Economic Review*, **101** (4), 1274–1311.
- FISCHMAN, J. B. (2013). Measuring inconsistency, indeterminacy, and error in adjudication. *American Law and Economics Review*, **16** (1), 40–85.
- FRANKEL, D. M., MORRIS, S. and PAUZNER, A. (2003). Equilibrium selection in global games with strategic complementarities. *Journal of Economic Theory*, **108** (1), 1–44.
- GAROUPA, N. and GINSBURG, T. (2009). Guarding the Guardians: Judicial Councils and Judicial Independence. *American Journal of Comparative Law*, **57** (1), 103–134.
- GLAESER, E. L. and SHLEIFER, A. (2002). Legal Origins. *Quarterly Journal of Economics*, **117** (4), 1193–1229.
- GREENE, J. D., CUSHMAN, F. A., STEWART, L. E., LOWENBERG, K., NYSTROM, L. E. and COHEN, J. D. (2009). Pushing moral buttons: The interaction between personal force and intention in moral judgment. *Cognition*, **111** (3), 364–371.
- GREIF, A. (1993). Contract enforceability and economic institutions in early trade: The maghribi traders’ coalition. *American Economic Review*, **83** (3), 525–548.

- HASTIE, T., TIBSHIRANI, R. and FRIEDMAN, J. (2008). *The Elements of Statistical Learning*. Springer.
- HELMKE, G. (2005). *Courts under Constraints: Judges, Generals and Presidents in Argentina*. Cambridge University Press.
- HENRICH, J. (2015). *The secret of our success: how culture is driving human evolution, domesticating our species, and making us smarter*. Princeton University Press.
- HOFFMAN, M., YOELI, E., DALKIRAN, N. A. and NOWAK, M. (2018). Categorical norms. Harvard University Working Paper.
- IMBENS, G. and KALYANARAMAN, K. (2012). Optimal bandwidth choice for the regression discontinuity estimator. *The Review of Economic Studies*, **79** (3), 933–959.
- KAHNEMAN, D., RITOV, I., SCHKADE, D., SHERMAN, S. J. and VARIAN, H. R. (1999). Economic preferences or attitude expressions?: An analysis of dollar responses to public issues. In *Elicitation of Preferences*, Springer, pp. 203–242.
- KARLAN, D. and LIST, J. A. (2006). *Does price matter in charitable giving? Evidence from a large-scale natural field experiment*. Tech. rep., National Bureau of Economic Research.
- KLAŠNJA, M. and TITIUNIK, R. (2017). The incumbency curse: Weak parties, term limits, and unfulfilled accountability. *American Political Science Review*, **111** (1), 129–148.
- KLING, J. (1999). *The Effect of Prison Sentence Length on the Subsequent Employment and Earnings of Criminal Defendants*. Princeton University, Discussion Papers in Economics, Discussion Paper No. 208.
- KLING, J. R. (2006). Incarceration length, employment, and earnings. *American Economic Review*, **96** (3), 863–876.
- KRAFT-TODD, G., YOELI, E., BHANOT, S. and RAND, D. (2015). Promoting cooperation in the field. *Current Opinion in Behavioral Sciences*, **3**, 96–101.
- LA PORTA, R., LOPEZ-DE SILANES, F., POP-ELECHES, C. and SHLEIFER, A. (2004). Judicial checks and balances. *Journal of Political Economy*.
- LAFFONT, J.-J. and MARTIMORT, D. (2009). *The theory of incentives: The principal-agent model*. Princeton university press.
- LEESON, P. T. (2007). An-arrgh-chy: The law and economics of pirate organization. *The Journal of Political Economy*, **115** (6), 1049–1094.
- LEVIN, J. (2003). Relational incentive contracts. *American Economic Review*, **93** (3), 835–857.
- MAILATH, G. J. and MORRIS, S. (2002). Repeated games with almost-public monitoring. *Journal of Economic Theory*, **102** (1), 189–228.
- and — (2006). Coordination failure in repeated games with almost-public monitoring. *Theoretical Economics*, **1** (3), 311–340.

- MARTINEAU, R. J. (1984). Frivolous appeals: The uncertain federal response. *Duke Law Journal*, **845**.
- MEHMOOD, S. (2019). Judiciary's achilles heel: Executive control via appointment power, unpublished.
- MONDERER, D. and SAMET, D. (1989). Approximating common knowledge with common beliefs. *Games and Economic Behavior*, **1** (2), 170–190.
- MONTESQUIEU, C. D. (1989). *Montesquieu: The spirit of the laws*. Cambridge University Press.
- MORRIS, S. (2002). Coordination, communication, and common knowledge: A retrospective on the electronic-mail game. *Oxford Review of Economic Policy*, **18** (4), 433–445.
- (2014). Coordination, timing and common knowledge. *Research in Economics*, **68** (4), 306–314.
- , ROB, R. and SHIN, H. S. (1995). p-dominance and belief potential. *Econometrica*, **63** (1), 145–157.
- and SHIN, H. S. (1997). Approximate common knowledge and co-ordination: Recent lessons from game theory. *Journal of Logic, Language and Information*, **6** (2), 171–190.
- and — (1998). Unique equilibrium in a model of self-fulfilling currency attacks. *American Economic Review*, **88** (3), 587–597.
- and — (2001). Global games: Theory and applications.
- NORTH, D. C. and WEINGAST, B. R. (1989). Constitution and Commitment: The Evolution of Institutions Governing Public Choice in Seventeenth-Century England. *Journal of Economic History*, **49** (4), 1–19.
- NOU (2002). *Dømmes av Likemenn [Judged by Peers]*. Ministry of Justice and Public Security, Norway (In Norwegian).
- NOWAK, M. A. (2006). Five rules for the evolution of cooperation. *Science*, **314** (5805), 1560–1563.
- and SIGMUND, K. (2005). Evolution of indirect reciprocity. *Nature*, **437** (7063), 1291–1298.
- OLKEN, B. A. and PANDE, R. (2012). Corruption in Developing Countries. *Annual Review of Economics*, **4**, 479–509.
- OSTROM, E. (2015). *Governing the commons*. Cambridge University Press.
- PALUMBO, G., GIUPPONI, G., NUNZIATA, L. and MORA-SANGUINETTI, J. S. (2013). Judicial performance and its determinants. *OECD Economic Policy Paper Series*.
- PINKER, S. (1997). *How the mind works*. Penguin UK.
- , NOWAK, M. A. and LEE, J. J. (2008). The logic of indirect speech. *Proceedings of the National Academy of Sciences*, **105** (3), 833–838.

- POBLETE-CAZENAVE, R. (2019). Crime and punishment: Do politicians in power receive special treatment in courts? evidence from india, unpublished.
- POSNER, R. A. (2000). Is the ninth circuit too large? a statistical study of judicial quality. *The Journal of Legal Studies*, **29** (2), 711–719.
- RAMSEYER, J. M. (2012). Talent matters: Judicial productivity and speed in japan. *International Review of Law and Economics*, **32** (1), 38–48.
- RAMSEYER, M. J. and RASMUSEN, E. B. (2001). Why are Japanese Judges so Conservative in Politically Charged Cases? *American Political Science Review*, **95** (2), 331–344.
- ROSALES-LÓPEZ, V. (2008). Economics of court performance: an empirical analysis. *European Journal of Law and Economics*, **25** (3), 231–251.
- RUBINSTEIN, A. (1989). The electronic mail game: Strategic behavior under "almost common knowledge". *The American Economic Review*, **97** (3), 385–391.
- SANCHEZ-MARTINEZ, C. A. (2017). Dismantling Institutions: Court Politicization and Discrimination in Public Employment Lawsuits, unpublished.
- SHAVELL, S. (1995). The appeals process as a means of error correction. *The Journal of Legal Studies*, **24** (2), 379–426.
- SHAYO, M. and ZUSSMAN, A. (2011). Judicial Ingroup Bias in the Shadow of Terrorism. *The Quarterly Journal of Economics*, **126** (3), 1447–1484.
- SPRANCA, M., MINSK, E. and BARON, J. (1991). Omission and commission in judgment and choice. *Journal of experimental social psychology*, **27** (1), 76–105.
- WEIBULL, J. W. (1997). *Evolutionary game theory*. MIT press.
- ZAFFALON, L. (2018). *A Política da Justiça: Blindar as Elites, Criminalizar os Pobres*. Editora Hucitec.

Appendix A

Appendix to Chapter 1

A.1 Parsing court documents

This section describes how we generate a data set of court case decisions by parsing the raw text in the Diários de Justiça. Parsing is a technology used by computers to read computer languages, and in computational linguistics to extract the meaning of natural language sentences. Since court documents are highly structured it turns out that using this technology works well in extracting key information from court documents. As an illustration of the method, assume, for simplicity, that the Diários de Justiça contains a sequence of paragraphs of the following form

00039966420148050110 - Ação de Improbidade Administrativa

Autor: Ministério Público Estadual

Réu: José Costa de Oliveira

Sentença: Julgo procedente

In order to build a parser to read this information one must first specify the logical structure, or the grammar, of the text. A common way to represent such grammars is in the so-called Backus-Naur form. The following grammar in Backus-Naur form represents the structure of the above example:

```
text := case*
```

```

case := number type plaintiff defendant decision
number := digit*
type := "– Acao de Improbidade Administrativa"
plaintiff := "Autor:" word*
defendant := "Reu:" word*
decision := "Sentenca:" ("Julgo procedente" | "Julgo improcedente")

```

The symbol "*" is read as "one or more" and "|" as "or". "digit" is any digit and "word" is any sequence of letters. Thus the grammar can be read as follows: The text consists of one or more cases, and a case consists of a case number, a lawsuit type, a plaintiff, a defendant, and a decision. Furthermore, a case number is a sequence of digits, a lawsuit type is either "Ação de Improbidade Administrativa" or "Ação Civil Pública", and so on. Once the grammar is specified, it can be inserted as input into parsing software which reads the text using the grammar, producing a parse tree. The parse tree can be transformed into a data set with the relevant information about each court case. Since each daily issue of a Diário contains up to several thousand pages we first extract the parts of each issue which deals with Ações de Improbidade using regular expressions before we run the parser, to save on computation time.

A.2 Incumbency disadvantage

That Brazil has an incumbency disadvantage for mayors has already been documented elsewhere (e.g. Klačnja and Titiunik (2017)). In this section we reproduce this result as well as providing evidence on the effect of marginally winning a mayoral election on longer term political outcomes. We also show results for city councilors. We estimate incumbency (dis)advantages using the same regression discontinuity specification as in the rest of the paper. In particular we use the bias-corrected estimator proposed by Calonico *et al.* (2014) with a local linear regression for the estimate and local quadratic regression for the bias-correction as our main specification. The specification of the local linear regression

is

$$y_{it} = \alpha + \beta E_{it} + \gamma WM_{it} + \delta E_{it} WM_i + \varepsilon_{ic}$$

where i is a politician and t is an election. The variable WM_{it} is the electoral win margin of the politician, and E_{it} is a dummy for whether the politician got elected. As outcomes y_{ic} we will look at future electoral outcomes. The first outcome we will study is the standard outcome in the incumbency disadvantage literature: Whether the politician gets elected to the position she was running for in the next election. We will also look at whether the politician gets elected to the position in the election after, eight years down the road. Finally, we will look at the total number of years that the politician is holding office in the future. This outcome counts any elected office, including state and federal offices. For the first outcome we will use the elections between 1998 and 2012, and for the last two outcomes we will look at the elections between 1998 and 2008. Outcomes always include electoral outcomes up until the 2018 general election. The results are presented in Table A.1. Column 1 of Panel A shows that whereas marginal losers in mayoral elections become the next mayor 38% of the times, this only happens in 22% of the times for marginal winners. Column 2 shows that the marginal loser also has a 13 percentage points higher probability of becoming the mayor in the subsequent election, eight years after. Finally, Column 3 shows that the expected future years in any elected office, starting the count from the next election, is 1.2 years larger for the marginal loser than for the marginal winner. Panel B shows that the effects are very similar if we run the same regression only among non-incumbents, which means that the result is not driven by term limits (incumbents who get reelected cannot run in the next election).¹ Panel C shows the same estimates for city councilors, which shows quite precise zeroes. For city councilors there is neither an incumbency advantage nor a disadvantage.

¹The reason the coefficients do not change much is that only 15% of candidates in close races are incumbents.

Table A.1: *Incumbency disadvantage*

Candidates for	Elected in next election	Elected in subseq. election	Future years in office
A: Mayor			
Elected	-0.16***	-0.13***	-1.2***
(se)	(0.0150)	(0.0095)	(0.1062)
N	24281	18093	18093
Bandwidth	0.072	0.154	0.096
Mean Marg. Loser	0.38	0.20	2.93
B: Mayor, non-incumbents			
Elected	-0.15***	-0.15***	-1.3***
(se)	(0.0168)	(0.0098)	(0.1129)
N	19105	14414	14414
Bandwidth	0.073	0.176	0.099
Mean Marg. Loser	0.41	0.21	3.12
C: City council			
Elected	-0.014	-0.016	0.11
(se)	(0.017)	(0.013)	(0.139)
N	14551	15096	15119
Bandwidth	0.088	0.115	0.091
Mean Marg. Loser	0.29	0.16	2.36

Notes: Standard incumbency (dis)advantage regressions comparing future political outcomes of marginal winners with marginal losers in Brazilian local elections. Column 1 uses the elections from 1998 through 2012, while Columns 2 and 3 exclude the 2012 election. The first two columns only consider whether the candidate is elected for the office they are running for in the next elections. 'Elected in subsequent election' is whether the candidate is elected to the office in the election eight years after the initial election. Future years in elected office include all elected offices, including state and federal positions. Regression discontinuity coefficients estimated using the bias-corrected estimator proposed by Calonico, Cattaneo, and Titiunik (2014) with a local linear regression for the estimate and local quadratic regression for the bias-correction. The running variable is the share of votes obtained in the election. No control variables. * $p \leq 0.1$; ** $p \leq 0.05$; *** $p \leq 0.01$.

Appendix B

Appendix to Chapter 2

B.1 Interpretation of the First-Stage Coefficient in Random Judge Designs

The empirical literature using random judge designs has commonly used leave-out mean judge characteristics as instruments for judge's case decisions (see, e.g., Bhuller *et al.* (forthcoming)). The first-stage equation in these analyses has the following form:

$$Z_{ik} = \gamma \bar{Z}_{j(ik)}^{-ik} + \theta X'_{ik} + \alpha_{ct} + \eta_{ik} \quad (\text{B.1})$$

where Z_{ik} is a case decision and the remaining variables are defined as in (2.1). Since the objective of such analyses is to identify the causal effects of case decision Z_{ik} on future case or defendant outcomes Y_{ik} , the instrument $\bar{Z}_{j(ik)}^{-ik}$ must satisfy the assumptions of exogeneity, monotonicity and exclusion for consistency. Although the objective of our paper is not to identify the causal effects of case decisions, but to assess how judge appeal tendency relates to future appeals and case or defendant outcomes, we do provide OLS regression estimates of (B.1) as part of the analysis presented in Section 2.4.1. To ease interpretation of the coefficient γ from such estimations, we thus provide some insights from the random judge design framework.

To simplify and without loss of generality, we can express equation (B.1) after residu-

alizing $\bar{Z}_{j(ik)}^{-ik}$ on characteristics X_{ik} and interacted court and case year fixed effects α_{ct} as follows:

$$\tilde{Z}_{ik} = \gamma \tilde{\bar{Z}}_{j(ik)}^{-ik} + \tilde{\eta}_{ik} \quad (\text{B.2})$$

where $(\tilde{Z}_{ik}, \tilde{\bar{Z}}_{j(ik)}^{-ik})$ are residualized variables and a regression of equation (B.2) recovers consistent estimates of γ under the same assumptions as would be needed for (2.1). Note now that since $\tilde{\bar{Z}}_{j(ik)}^{-ik}$, being a leave-out measure, is essentially very close to the judge-specific mean of the outcome $\tilde{Z}_{ik}$, in *large samples* (as $N_j \rightarrow \infty$) the regression coefficient estimate of γ shall approach 1. In small samples (with few cases per judge, i.e., N_j is small), leave-out sample appeal rate can be a noisy measure of the true judge appeal rate:

$$\tilde{\bar{Z}}_{j(ik)}^{-ik} = \tilde{\bar{Z}}_j + u_{ij}$$

where u_{ij} is a white-noise term with mean 0, which is also uncorrelated with $\tilde{\bar{Z}}_j$ and $\tilde{Z}_{ik}$ by the virtue of random assignment of cases to judges. In other words the measurement error in $\tilde{\bar{Z}}_{j(ik)}^{-ik}$ is *classical*, implying that the estimated $\hat{\gamma}$ we get from equation (B.2) when using $\tilde{\bar{Z}}_{j(ik)}^{-ik}$ as a measure of true judge appeal rate $\tilde{\bar{Z}}_j$ will have the following attenuation bias:

$$\hat{\gamma} = \gamma \frac{\sigma_{\tilde{\bar{Z}}_j}^2}{\sigma_{\tilde{\bar{Z}}_{j(ik)}^{-ik}}^2}$$

Since the true $\gamma = 1$, this implies:

$$\sigma_{\tilde{\bar{Z}}_j}^2 = \hat{\gamma} \sigma_{\tilde{\bar{Z}}_{j(ik)}^{-ik}}^2 \quad (\text{B.3})$$

Thus, one can interpret $\hat{\gamma}$ as the share of the variance in leave-out judge appeal tendency (residualized) that can be attributed to systematic variation in appeal rates across judges as opposed to noise.

B.2 Proofs for Propositions 1–3

Proofs

Proof. (Proposition 1). Assume

$$Z_{ik} = bS_{ik} + \xi_{ik}$$

$$Z_{ik} = \beta \bar{S}_{j(ik)}^{-ik} + \gamma \bar{Z}_{j(ik)}^{-ik} + \varepsilon_{ij}$$

with ξ_{ik} iid and $\varepsilon_{ij} \perp \bar{S}_{j(ik)}^{-ik}, \bar{Z}_{j(ik)}^{-ik}$. By the last assumption we have

$$\text{Cov} \left(\bar{Z}_{j(ik)}^{-ik}, \varepsilon_{ij} \right) = 0$$

$$\text{Cov} \left(b\bar{S}_{j(ik)}^{-ik}, \varepsilon_{ij} \right) = 0$$

Using

$$\varepsilon_{ij} = \xi_{ij} + bS_{ik} - \beta \bar{S}_{j(ik)}^{-ik} - \gamma \bar{Z}_{j(ik)}^{-ik}$$

we then get

$$\begin{aligned} \text{Cov} \left(bS_{ik}, \bar{Z}_{j(ik)}^{-ik} \right) &= \text{Cov} \left(\beta \bar{S}_{j(ik)}^{-ik}, \bar{Z}_{j(ik)}^{-ik} \right) + \gamma \text{Var} \left(\bar{Z}_{j(ik)}^{-ik} \right) \\ \text{Cov} \left(bS_{ik}, \bar{S}_{j(ik)}^{-ik} \right) &= \text{Cov} \left(\beta \bar{S}_{j(ik)}^{-ik}, \bar{S}_{j(ik)}^{-ik} \right) + \gamma \text{Cov} \left(\bar{Z}_{j(ik)}^{-ik}, \bar{S}_{j(ik)}^{-ik} \right) \end{aligned}$$

Substituting for $\bar{Z}_{j(ik)}^{-ik}$ and multiplying the last equation with b gives

$$\begin{aligned} \text{Cov} \left(bS_{ik}, b\bar{S}_{j(ik)}^{-ik} \right) &= \text{Cov} \left(\beta \bar{S}_{j(ik)}^{-ik}, b\bar{S}_{j(ik)}^{-ik} \right) + \gamma \text{Var} \left(b\bar{S}_{j(ik)}^{-ik} \right) + \gamma \text{Var} \left(\varepsilon_{ij} \right) \\ \text{Cov} \left(bS_{ik}, b\bar{S}_{j(ik)}^{-ik} \right) &= \text{Cov} \left(\beta \bar{S}_{j(ik)}^{-ik}, b\bar{S}_{j(ik)}^{-ik} \right) + \gamma \text{Cov} \left(b\bar{S}_{j(ik)}^{-ik}, b\bar{S}_{j(ik)}^{-ik} \right) \end{aligned}$$

If we subtract the second equation from the first we get

$$\gamma \text{Var} \left(\varepsilon_{ij} \right) = 0 \Rightarrow \gamma = 0$$

□

Proof. (Proposition 2). Assume that

$$\bar{Z}_j = \mu \bar{S}_j + v_j$$

where v_j is an iid error term. This gives

$$\mu = \frac{\text{Cov}(\bar{Z}_j, \bar{S}_j)}{\text{Var}(\bar{S}_j)}$$

Thus

$$\mu^2 \text{Var}(\bar{S}_j) = \frac{\text{Cov}(\bar{Z}_j, \bar{S}_j)^2}{\text{Var}(\bar{S}_j)}$$

Since $\bar{S}_{ij}$ and $\bar{S}_{j(ik)}^{-ik}$ are independent measures of $\bar{S}_j$ we have

$$\text{plim Cov}(\bar{S}_{ij}, \bar{S}_{j(ik)}^{-ik}) = \text{Var}(\bar{S}_j)$$

Similarly

$$\text{plim Cov}(\bar{Z}_{ij}, \bar{S}_{j(ik)}^{-ik}) = \text{Cov}(\bar{Z}_j, \bar{S}_j)$$

□

Variance Decomposition with Multi-dimensional Judge Stringency.

The following extends Proposition 2 to settings where the appeal tendency of a judge might be independently determined by several dimensions of judge stringencies. The variances and covariances can be estimated in the same way as in the proof of Proposition 2.

Proposition 3. Assume that $\bar{S}_j$ is a vector of judge stringency measures and that

$$\bar{Z}_j = \mu^T \bar{S}_j + v_j$$

Then the variance in $\bar{Z}_j$ which can be attributed to differences in $\bar{S}_j$ is

$$\text{Cov}(\bar{S}_j, \bar{Z}_j)^T \text{Var}(\bar{S}_j)^{-1} \text{Cov}(\bar{S}_j, \bar{Z}_j)$$

Proof. We have that

$$\mu = \text{Var}(\bar{S}_j)^{-1} \text{Cov}(\bar{S}_j, \bar{Z}_j)$$

This gives

$$\text{Var}(\mu^T \bar{S}_j) = \mu^T \text{Var}(\bar{S}_j) \mu = \text{Cov}(\bar{S}_j, \bar{Z}_j)^T \text{Var}(\bar{S}_j)^{-1} \text{Cov}(\bar{S}_j, \bar{Z}_j)$$

□

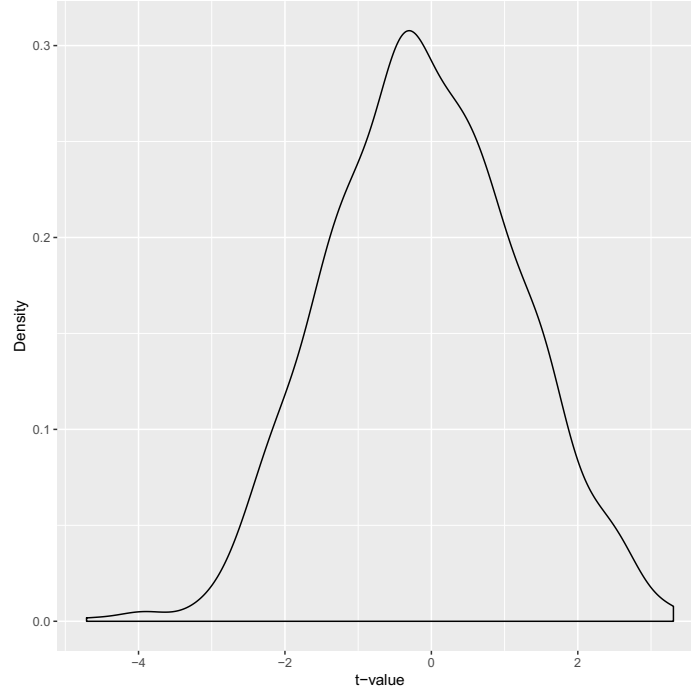
B.3 Monte Carlo Simulations

In order to check the small sample properties of the test of whether all variation in appeal rates across judges is driven by differences in judge stringencies described in Section 2.3.4 we run the following Monte Carlo simulation. First, we randomly draw judge stringencies of 539 judges from a normal distribution with mean 0.4 and standard deviation 0.05. Judges who are assigned stringencies lower than 0 or greater than 1 are assigned stringencies of 0 and 1, respectively. We then draw 122 cases per judge and determine incarceration iid across cases with probability of incarceration equal the stringency parameter of the judge. Subsequently, we draw whether each case is appealed with probability 0.5 when there is an incarceration, and 0.1 when not. The numbers are chosen to match moments from our main estimation sample. Using the simulated data we then run the following regression

$$Z_{ik} = \beta \bar{S}_{j(ik)}^{-ik} + \gamma \bar{Z}_{j(ik)}^{-ik} + \varepsilon_{ij}$$

We run 1000 simulations and report the distribution of the t -value of $\hat{\gamma}$ in Figure B.1.

Figure B.1: *The Empirical Distribution of t-values from Monte Carlo Simulations.*



Note: Monte Carlo simulations showing the distribution of t -values of $\hat{\gamma}$ from running the regression $Z_{ik} = \beta \bar{S}_{j(ik)}^{-ik} + \gamma \bar{Z}_{j(ik)}^{-ik} + \varepsilon_{ij}$ on simulated data assuming that the only determinant of appeal is the harshness of a sentence. Number of judges and cases per judge the same as in the real data. Parameters determining appeal rates and incarceration stringencies across judges chosen to match sample moments. 1000 simulations.

B.4 Measuring Judge Disagreement

Let j be a judge and let ik represent a randomly drawn court case k involving defendant i . Denote by X'_{ik} a vector of court case characteristics and by D_{ikj} whether the judge incarcerates. Assume that judge utility is

$$U_{ikj} = \begin{cases} \lambda_j X'_{ik} + \kappa_{ikj} & \text{if } D_{ikj} = 1 \\ 0 & \text{if } D_{ikj} = 0 \end{cases}$$

where κ_{ikj} is an extreme value distributed error term across ik . We define the *disagreement* of the judge j with judge l as

$$d \equiv \Pr [D_{ikj} \neq D_{ikl}]$$

Assume that

$$\kappa_{ik1} = \kappa_{ik2}$$

Theorem 1. *Judge Disagreement is identified by*

$$d = E_{ik} \left[\frac{|e^{-\lambda_1 X'_{ik}} - e^{-\lambda_2 X'_{ik}}|}{(1 + e^{-\lambda_1 X'_{ik}})(1 + e^{-\lambda_2 X'_{ik}})} \right]$$

Proof. (Theorem 1). The logistical distribution of κ_{ikj} implies

$$P [D_{ikj} = 1 | X'_{ik}] = \frac{1}{1 + e^{-\lambda_j X'_{ik}}}$$

Assume that $P [D_{ik1} = 1 | X'_{ik}] \geq P [D_{ik2} = 1 | X'_{ik}]$. This implies that judge 1 will incarcerate whenever 2 incarcerates, i.e., *conditional monotonicity*. Their disagreement on cases with observable characteristics X'_{ik} is thus

$$\frac{1}{1 + e^{-\lambda_1 X'_{ik}}} - \frac{1}{1 + e^{-\lambda_2 X'_{ik}}} = \frac{e^{-\lambda_1 X'_{ik}} - e^{-\lambda_2 X'_{ik}}}{(1 + e^{-\lambda_1 X'_{ik}})(1 + e^{-\lambda_2 X'_{ik}})}$$

The result is obtained by aggregating across all cases.

□

Estimation of Judge Disagreement.

In order to estimate λ_j we encounter the curse of dimensionality: We have access to many observable characteristics X'_{ik} , but observe only on average 250 decisions per judge. In order to reduce the dimensionality we rely on regularization. In particular, for each judge j , we maximize the following penalized log likelihood

$$\max_{\lambda_j} \left\{ \sum_{ik=1}^{N_j} \left[D_{ikj} P(X'_{ik}; \lambda_j) + (1 - D_{ikj}) \left[1 - P(X'_{ik}; \lambda_j) \right] - \psi \sum_{m=1}^M |\lambda_{jm}| \right] \right\}$$

where $\lambda_j = \{\lambda_{j0}, \lambda_{j1}, \dots, \lambda_{jM}\}$ and

$$P(X'_{ik}; \lambda_j) \equiv \frac{1}{1 + e^{-\lambda_j X'_{ik}}}$$

This is a variant of the Lasso regression for the logistic regression where the parameter ψ serves to penalize large coefficients, and tends to reduce the number of variables with a positive λ_{jm} .¹ We choose the value of ψ by 5-fold *cross-validation*: The data is split into 5 “folds”. On each fold we compute the mean squared prediction error of the model when it is estimated using the remaining 4 folds. We choose the ψ which gives the lowest average mean squared prediction error.

¹For details see, e.g., Hastie *et al.* (2008), chapter 4.4.4.

Appendix C

Appendix to Chapter 3

C.1 Characterizing Which Signals Can Trigger Sanctions

Theorem 1. Consider the coordination game with signals $\Gamma = (G, \Omega, S, \mu, F)$. Consider any $S_1^*, S_2^* \subseteq S$ and define the strategy $\sigma_i^*(s_i) = A$ iff $s_i \in S_i^*$. The strategy profile (σ_1^*, σ_2^*) is a B.N.E. of Γ iff for $i = 1, 2$:

$$s_i \in S_i^* \Rightarrow \text{player } i \text{ } \bar{p}\text{-believes } \mathbf{S}_{-i}^* \quad (\text{C.1})$$

$$s_i \notin S_i^* \Rightarrow \text{player } i \text{ } (1 - \bar{p})\text{-believes } \neg \mathbf{S}_{-i}^* \quad (\text{C.2})$$

Proof. This follows immediately from the definitions of $\bar{p}$, p -beliefs, and best responses.

Let $S_1^*, S_2^* \subseteq S$, and $\sigma_i^*(s_i) = A$ iff $s_i \in S_i^*$.

If direction: Suppose for $i = 1, 2$, $s_i \in S_i^* \Rightarrow \text{player } i \text{ } \bar{p}\text{-believes } \mathbf{S}_{-i}^*$, and $s_i \notin S_i^* \Rightarrow \text{player } i \text{ } (1 - \bar{p})\text{-believes } \neg \mathbf{S}_{-i}^*$ (A1). We want to show (σ_1^*, σ_2^*) is a B.N.E.. I.e. choose an i , an $s_i \in S_i^*$, and an $s'_i \notin S_i^*$. We want to show A is a best response to the belief that the other player will play A with probability $P(\mathbf{S}_{-i}^* | \mathbf{s}_i)$, and B is a best response to the belief that the other player will play B with probability $P(\mathbf{S}_{-i}^* | \mathbf{s}'_i)$.

By assumption (A1), $P(\mathbf{S}_{-i}^* | \mathbf{s}_i) \geq \bar{p}$, so, by the definition of risk dominance, player i 's

best response given s_i is to play A . Also by assumption (A1), $P(\mathbf{S}_{-i}^* | \mathbf{s}_i') \leq \bar{p}$. Then, again, by the definition of risk dominance, player i 's best response given s_i' is to play B . $\square$

Only if direction: Suppose (σ_1^*, σ_2^*) is a B.N.E (A2). We want to show for $i = 1, 2$, $s_i \in S_i^* \Rightarrow$ player i $\bar{p}$ -believes $\mathbf{S}_{-i}^*$, and $s_i \notin S_i^* \Rightarrow$ player i $(1 - \bar{p})$ -believes $\neg \mathbf{S}_{-i}^*$. Each of these follows again from the definition of risk dominance and best response. $\square$

C.2 The Connection With Higher Order Beliefs

Corollary 1. Consider the coordination game with signals $\Gamma = (G, \Omega, S, \mu, F)$. Consider any $S_1^*, S_2^* \subseteq S$ and define the strategy $\sigma_i^*(s_i) = A$ iff $s_i \in S_i^*$. Let $\mathbf{E}^* = \Omega \times S_1^* \times S_2^*$ and $\mathbf{F}^* = \Omega \times \neg S_1^* \times \neg S_2^*$. Then, the strategy profile (σ_1^*, σ_2^*) is a B.N.E. of Γ iff $\mathbf{E}^*$ is $\bar{p}$ -evident and $\mathbf{F}^*$ is $(1 - \bar{p})$ -evident.

Proof. This follows immediately from Thm. 1, the definition of $\bar{p}$ -evident, and the construction of $\mathbf{E}^*$ and $\mathbf{F}^*$.

Notice that whenever $\mathbf{E}^*$ occurs, player i gets some signal $s_i' \in S_i^*$, at which point player i $\bar{p}$ -believes $\mathbf{E}^*$ with exactly the probability utilized in Thm. 1, $P(\mathbf{S}_{-i}^* | \mathbf{s}_i')$. Thus, $\mathbf{E}^*$ is $\bar{p}$ -evident iff for all i , i $\bar{p}$ -believes S_{-i}^* for all $s_i \in S_i^*$.

Likewise, whenever $\mathbf{F}^*$ occurs, player i gets some signal $s_i'' \in \neg S_i^*$. At which point player i $\bar{p}$ -believes $\mathbf{F}^*$ with exactly the probability $P(\neg \mathbf{S}_{-i}^* | \mathbf{s}_i'')$. Thus, $\mathbf{F}^*$ is $(1 - \bar{p})$ -evident iff for all i , player i $1 - \bar{p}$ -believes $\neg \mathbf{S}_{-i}^*$ for all $s_i \in \neg S_i^*$.

Hence, $\mathbf{E}^*$ is $\bar{p}$ -evident and $\mathbf{F}^*$ is $(1 - \bar{p})$ -evident, iff (σ_1^*, σ_2^*) is a B.N.E.. $\square$

C.3 The Connection With Deterrence

Theorem 2. Action a^* can be enforced if and only if there exists $S_1^*, S_2^* \subseteq S$ such that:

$$\text{for } i = 1, 2 : s_i \in S_i^* \Leftrightarrow i \text{ } \bar{p}\text{-believes } \mathbf{S}_{-i}^* \text{ given } \mu_{a^*} \quad (\text{C.3})$$

$$a^* \in \arg \max_{a \in A} V(a, \sigma_1^*, \sigma_2^*), \text{ where for } i \in \{1, 2\}, \sigma_i^*(s_i) = A \text{ iff } s_i \in S_i^* \quad (\text{C.4})$$

Proof. Suppose conditions 3.3 and 3.4 are satisfied. That is, suppose for $i = 1, 2 : s_i \in S_i^* \Leftrightarrow i \text{ } \bar{p}\text{-believes } S_{-i}^* \text{ given } \mu_{a^*} \text{ and } a^* \in \arg \max_{a \in A} V(a, \sigma_1^*, \sigma_2^*)$, where for $i \in \{1, 2\}, \sigma_i^*(s_i) = A \text{ iff } s_i \in S_i^*$.

Consider the strategy profile $(a^*, \sigma_1^*, \sigma_2^*)$. Notice that the agent cannot benefit from deviating, by condition 3.4. Likewise, the enforcers cannot benefit from deviating, given each's strategy and the agent's, based on the same logic as in the proof of Thm. 1. Note that the use of μ_{a^*} instead of μ incorporates the enforcers' expectations of the agent's behavior and the σ_i^* s used in V incorporate the agent's expectation of the enforcers' behavior.

To see the other direction, suppose that conditions 3.3 and 3.4 cannot simultaneously be satisfied for any S_1^*, S_2^* . Consider some S_1^+ and S_2^+ , such that a^* is a best response given each player $i \in \{1, 2\}$ plays $\sigma_i^+(s_i) = A \text{ iff } s_i \in S_i^+$. Then, assumption 3.3 must not be satisfied, if S_i^* is replaced with S_i^+ for $i \in \{1, 2\}$. But, this implies, by the same logic as in Thm. 1, that for an enforcer i , given her expectations that the agent plays a^* and enforcer $-i$ plays A on S_{-i}^+ , there is some signal s_i such that she will benefit from deviating. $\square$

C.4 Categorical Norms: An Example With Uniform Distributions of States and Signals

Proposition 1. *For any $\underline{s} \in S$, the threshold norm at $\underline{s}$ is a B.N.E. iff $\bar{p} = .5$.*

Proof. Notice that for a threshold norm at $\underline{s}$ to be a B.N.E., the players need to be indifferent between A and B at $\underline{s}$, given their belief that the other plays A for all signals above $\underline{s}$ and B for all signals below $\underline{s}$. But, at $\underline{s}$, player i believes player $-i$ gets a signal above $\underline{s}$ with probability exactly .5, and below $\underline{s}$ with probability exactly .5, and thus player i believes player $-i$ plays A with probability exactly .5 and B with probability exactly .5. So, player i will only be indifferent if $p = .5$. $\square$

C.5 Categorical Norms: A Result for Arbitrary Distributions of States and Signals

Theorem 3. For any $\underline{s} \in S$, a threshold norm at $\underline{s}$ is a B.N.E. for $i \in \{1, 2\}$ iff:

$$\sup_{s_i < \underline{s}} P(s_{-i} < \underline{s} | s_i) \geq 1 - \bar{p} \geq \inf_{s_i > \underline{s}} P(s_{-i} < \underline{s} | s_i)$$

Proof. See Hoffman *et al.* (2018). The intuition is as follows. Because of the MLRP, the signals at which it is most tempting to deviate are selected by sup and inf. Whether player i would want to deviate at these points depends on whether the probability they place on the other player's signal being on the same side of the threshold as their signal is at least $\bar{p}$ and $1 - \bar{p}$, respectively. $\square$

C.6 The Role of Observability and Correlation

Proposition 2. Let $S_1^* = S_2^* = \{1\}$ and define the strategy $\sigma_i^*(s_i) = A$ iff $s_i \in S_i^*$. The strategy profile (σ_1^*, σ_2^*) is a B.N.E. of Γ iff:

$$1 - \epsilon(1 - \rho) \geq \bar{p} \tag{C.5}$$

$$\frac{\mu_1 \epsilon(1 - \epsilon)(1 - \rho)}{\mu_1 \epsilon + (1 - \mu_1)} \leq \bar{p} \tag{C.6}$$

Proof. Prop. 2 follows from Thm. 1, the current setup, and the derivation of ρ from r .

In particular, condition 3.5 corresponds to condition 3.1 in Thm. 1. When player i gets the signal 1, for player i to have no incentive to deviate from playing A as prescribed, player i needs to $\bar{p}$ -believe that player $-i$ received signal 1 and is thus playing A . But, the probability that player $-i$ received 1 given than player i receives 1 is exactly r . By the logic presented in footnote 3 relating r to ρ , and after some algebra, $r = 1 - \epsilon(1 - \rho)$.

Likewise, condition 3.6 corresponds to condition 3.2. When player i gets the signal 0,

for player i to have no incentive to deviate from playing B as prescribed, player i needs to $(1 - \bar{p})$ -believe that player $-i$ received signal 0 and is thus playing B . We now calculate this probability:

$$\begin{aligned}
P(s_{-i} = 1 | s_i = 0) &= \frac{P(s_{-i} = 1, s_i = 0)}{P(s_i = 0)} \\
&= \frac{P(s_{-i} = 1, s_i = 0 | \omega = 1) \mu_1 + P(s_{-i} = 1, s_i = 0 | \omega = 0) (1 - \mu_1)}{P(s_i = 0 | \omega = 1) \mu_1 + P(s_i = 0 | \omega = 0) (1 - \mu_1)} \\
&= \frac{P(s_{-i} = 1, s_i = 0 | \omega = 1) \mu_1 + 0}{\epsilon \mu_1 + 1 - \mu_1} \tag{C.7}
\end{aligned}$$

$$\begin{aligned}
&= \frac{(1 - \epsilon - P(s_{-i} = 1, s_i = 1 | \omega = 1)) \mu_1}{\epsilon \mu_1 + 1 - \mu_1} \\
&= \frac{(1 - \epsilon - P(s_{-i} = 1 | s_i = 1) (1 - \epsilon)) \mu_1}{\epsilon \mu_1 + 1 - \mu_1} \tag{C.8} \\
&= \frac{(1 - \epsilon - (\rho \epsilon + (1 - \epsilon)) (1 - \epsilon)) \mu_1}{\epsilon \mu_1 + 1 - \mu_1} \\
&= \frac{(1 - \rho) \epsilon (1 - \epsilon) \mu_1}{\epsilon \mu_1 + 1 - \mu_1}
\end{aligned}$$

At line C.7, we use the following equality, $P(s_{-i} = 1, s_i = 1 | \omega = 1) + P(s_{-i} = 1, s_i = 0 | \omega = 1) = P(s_{-i} = 1 | \omega = 1) = 1 - \epsilon$. At line C.8 we use the definition of r , $P(s_{-i} = 1 | s_i = 1) = r$, and the relationship between r and ϵ, ρ calculated earlier, $r = 1 - \epsilon(1 - \rho)$. $\square$