CERGE
Center for Economics Research and Graduate Education
Charles University Prague

Essays in Applied Economics

Michal Šoltés

Dissertation

Prague, October 2021

Michal Šoltés

Essays in Applied Economics

Dissertation

Prague, October 2021

Dissertation Committee

Filip Matějka (CERGE-EI; Chair)

Michal Bauer (CERGE-EI)

Fabio Michelucci (Università Ca' Foscari)

Nikolas Mittag (CERGE-EI)

Referees

Christoph Engel (Max Planck Institute)

Francesco Drago (University of Catania)

Table of Contents

Abstract

In the first chapter, we study the role of political parties in the selection of candidates in Czech municipal elections. Using over 20,000 electoral slates, we document that political parties rank candidates on the slates according to their valence and intra-party value. Valence, measured by education level, captures the public value of the candidates, while intra-party value, measured by political donations and membership, represents the value of the candidate to the party. The patterns we observe are consistent with market mechanisms between candidates and party leaders where the party leaders benefit from the valence and intra-party value of candidates and offer slate positions (i.e. the probability of winning a mandate) in exchange. We show that candidates with high valence and those who possess more intra-party value are placed in better ranked positions, despite the fact that candidates with more intra-party value, conditional on observables, tend to receive relatively fewer votes than candidates with low intra-party value. We also show that as a party expects to hold more council seats, the share of their candidates with higher intra-party value increases.

In the second chapter, we study the implications of a structure of the criminal justice system on sentencing decisions. To limit sentencing disparities, offenses are typically divided into subsections with specific sentencing ranges. The classification into corresponding subsections often depends on exceeding a given quantity threshold, such as drug amount. We study the consequences of these quantity thresholds on sentencing decisions and argue that the threshold effect can be decomposed into two opposing mechanisms: the severity mechanism and the reference one. An experiment with Czech prosecutors shows that thresholds cause an enormous increase in harshness of sentencing, leading to sentencing disparities. We further introduce empirical measures of (in)justice and quantify the consequences of quantity thresholds on the probability of imposing a just sentence.

In the third chapter, I explore consequences of publishing inconvenient information about the performance of public institutions. To understand how citizens would respond to such information, I conducted a survey experiment in which respondents were informed about sentencing disparity in the Czech Republic caused by different practices of imposing sentences among judges, i.e. information that likely questions the competence of the criminal justice system to deliver on its responsibility. The results suggest that such information does not lead to distrust and avoidance of the formal judicial system. Instead, the treated

respondents became more likely to sign a petition that invites politicians to address the underlying issue, and respondents found fairness of the judicial system as a more important policy issue. I found sizeable heterogeneity in the treatment effect. The increase in the willingness to sign a petition was driven by mothers, who are arguably more sensitive to the particular treatment information in the presented case of a failure to pay alimony.

Abstrakt

V první kapitole studujeme vliv politických stran při výběru kandidátů v komunálních volbách v České republice. Za použití dat z více než 20 000 kandidátních listin ukazujeme, že politické strany řadí kandidáty na kandidátku systematicky podle jejich kvalit a stranické hodnoty. Kvalita kandidátů, měřena nejvyšším dosaženým vzděláním, zachycuje hodnotu kandidáta pro voliče, zatímco stranická hodnota, měřená politickými dary nebo členstvím ve straně, reprezentuje vnitřní hodnotu kandidáta pro stranu. Pozorované chování je konzistentní s tržní situací, kdy kandidáti nabízejí lídrovi kandidátní listiny jejich kvalitu a stranickou hodnotu výměnou za lepší postavení na kandidátní listině (a tedy i vyšší pravděpodobností zvolení). Na datech ukazujeme, že kvalitnější kandidáti a kandidáti s vyšší stranickou hodnotou jsou řazeni na lepších místech na kandidátce, a to navzdory tomu, že kandidáti s vyšší stranickou hodnotou získávají ve volbách zpravidla méně hlasů než kandidáti s nízkou stranickou hodnotou, *ceteris paribus*. Dále ukazujeme, že pokud strana očekává lepší volební výsledek, a tedy i více zastupitelských míst, vzroste na její kandidátní listině podíl kandidátů s vyšší stranickou hodnotou.

V druhé kapitole studujeme dopady struktury zvláštní části trestního zákoníku na ukládání trestů. Ve snaze omezit rozdíly v uložených trestech jsou trestné činy zpravidla rozdělené do několika odstavců dle závažnosti, přičemž pro každý odstavec je určena vlastní trestní sazba. Ke stanovení závažnosti zákonodárce často používá striktní hodnoty měřitelné veličiny, například způsobenou škodu. V kapitole studujeme důsledky rozdělení trestných činů na odstavce na ukládání trestů. Vliv takové hranice lze rozložit na dva mechanismy působící v opačném směru: mechanismus přísnosti a mechanismus srovnání. V experimentu se státními zástupci ukazujeme, že striktní dělení trestného činu na odstavce dramaticky zvýšilo tvrdost navrženého trestu, a způsobilo tak významné rozdíly v uložených trestech. Dále v kapitole představujeme nový způsob empirického měření spravedlnosti uložených trestů a vyčíslujeme dopady striktního dělení odstavců na pravděpodobnost udělení spravedlivého trestu.

Ve třetí kapitole se zabývám důsledky zveřejnění nevyhovujících informací o výkonu veřejných institucí. V rámci dotazníkového experimentu byli respondenti informováni o rozdílech v uložených trestech v České republice způsobených rozdílným přístupem soudců k ukládání trestů. Respondentům tak byly poskytnuty informace, které mohou vyvolat pochyby o schopnosti trestní justice dostát své zodpovědnosti. Výsledky ukazují,

že taková informace nezpůsobila nedůvěru ani nezvýšila snahu veřejnosti vyhnout se formálnímu soudnímu systému. Poskytnutí informace místo toho zvýšilo pravděpodobnost, že respondent bude ochotný podepsat petici, která vyzývá politiky k řešení daného problému. Poskytnuté informace dále zvýšily vnímanou důležitost spravedlnosti právního systému. Vliv informací je heterogenní. Nárůst ochoty podepsat petici je patrný především mezi matkami, které jsou zřejmě více citlivé na konkrétní představený trestný čin neplacení alimentů.

Acknowledgments

I am grateful to many people who have helped me throughout my studies and supported me in writing the dissertation. I am greatly indebted to Filip Matějka. He has helped me immensely every step of the way, and for that, I will be forever grateful. It is not an overstatement to say that without him, this thesis would not exist.

I have benefited from many discussions and lectures. In particular, Michal Bauer, Nikolas Mittag, Štěpán Jurajda, and Jan Zápal have shaped my economic thinking significantly. When facing questions and problems in economics, I often ask myself how they would approach it. The imaginary recommendation is, surprisingly often, to take a step back and write the problem down piece by piece.

Jakub Drápal, Vladimír Novák, and Miroslav Palanský deserve special credit for their patience and time. I do not remember any midnight research ideas or random thought about the academic world that I would not share with (at least) one of them. Luckily for me, they never ignored the idea, or me for that matter, for more than a couple of days. Furthermore, they seem to tolerate and sometimes even understand my occasional complaints about life (of a Ph.D. student). These (academic) friendships and collaborations are invaluable and an excellent benefit of academia.

I am also grateful to several classmates who helped me through the demanding years. Without their help, the coursework would have been noticeably more challenging.

Recently, I have started several research projects with internationally recognized economists. I am excited about the opportunity and pleased to see that CERGE-EI education does open many research possibilities. For that, I will always be indebted to the CERGE-EI community, faculty, and staff for that.

I extend my thanks to my family and friends for their support. I could not devote as much of my time and energy to you as I would have like.

I would especially like to thank my wife, who has been infinitely patient with me. Thank you for teaching me that there are more important things than research and dissertation work. I know it is a slow process, but I will get that eventually. I promise.

Financial support from Charles University, Charles University, GAUK project No. 218218, from the European Research Council under the European Union's Horizon 2020 research and innovation program (grant agreement No 678081), from H2020-MSCA-RISE project GEMCLIME-2020 (grant agreement no 681228), and from Nadace Nadání Josefa, Marie a Zdeňky Hlávkových is gratefully acknowledged.

Czech Republic, Prague
October 2021

Michal Šoltés

Introduction

This thesis consists of three chapters, each devoted to a separate question. In the first chapter, Klára Svitáková and I study the role of political parties in the selection of candidates in Czech municipal elections. The quality of elected candidates is a key aspect and a necessary condition for good governance. Therefore, it is essential to understand who becomes elected and, as the first step, who runs for office. In many systems, it is political parties that determine who will run for office. Furthermore, in electoral systems, in which the initial slate position of candidates is highly informative about winning a mandate, political parties also affect the probability of winning a mandate. As a result, political parties have great power over elected politicians, and it is thus essential to understand how political parties form their slates and what drives the candidates' positions.

We leverage a convenient institutional design and study how the intra-party value of candidates interacts with their slate positions in municipal elections in the Czech Republic. In particular, using over 20,000 electoral slates, we document that political parties rank candidates on the slates systematically according to their valence and intra-party value. Valence, measured by education level, captures the public value of the candidates. In contrast, intra-party value, measured by political donations and membership, represents the value of the candidate to the party. The patterns we observe are consistent with market mechanisms between candidates and party leaders. The party leaders benefit from the valence and intra-party value of candidates and offer slate positions (i.e., the probability of winning a mandate) in exchange. We show that candidates with high

valence and those who possess more intra-party value are placed in better-ranked positions, although candidates with more intra-party value, conditional on observables, tend to receive relatively fewer votes than candidates with low intra-party value. We also show that as a party expects to hold more council seats, the share of their candidates with higher intra-party value increases. Overall, we provide strong evidence that political parties skew political representation based on a quid pro quo relationship with the candidates.

In the second chapter, Jakub Drápal and I study sentencing decisions. Many scholars in criminology, the economics of crime, and law have documented sizeable sentencing disparities resulting from judges' and prosecutors' different practices. Depending on the judge and prosecutor assigned, the same case can be sentenced differently. We add to this literature by describing a new source of sentencing disparities: Offense subsection with specific sentencing ranges.

Many criminal justice systems divide offenses into subsections with specific sentencing ranges. The classification into corresponding subsections often depends on exceeding a given quantity threshold, such as the amount of a drug or damage caused. We study the consequences of these quantity thresholds on sentencing decisions and argue that the threshold effect can be decomposed into two opposing mechanisms: the severity mechanism and the reference one. The underlying theory is also informative for potential policies that would change the structure of the criminal code. To provide empirical evidence, we conduct a Rachlinski-style online experiment with 200 Czech prosecutors. The experiment shows that thresholds cause an enormous increase in harshness in sentencing, leading to sentencing disparities. Finally, we introduce empirical measures of justice and injustice and quantify the consequences of quantity thresholds on the probability of just sentences.

In the third chapter, I explore the consequences of publishing inconvenient information about the performance of public institutions. Many public institutions have been found reluctant to publish inconvenient information that may reveal incompetence and/or systematic failure of responsible authorities. To avoid publishing, institutions often rely on the argument that publishing sensitive information would cause distrust and have adverse consequences.

To understand how citizens would respond to such information. I conducted a survey experiment in which respondents were informed about the sentencing disparity in the Czech Republic caused by the different practices of imposing sentences among judges,

i.e., information that likely questions the competence of the judicial system. The results suggest that such information does not lead to distrust and avoidance of the formal judicial system. Instead, the treated respondents became more likely to sign a petition that invites politicians to address the issue, and respondents found fairness of the judicial system as a more important policy issue. I found sizeable heterogeneity in the treatment effect. The increase in the willingness to sign a petition was driven by mothers, who are arguably more sensitive to the particular treatment information that concerned a failure to pay alimony.

Chapter 1

Ranking of Candidates on Slates: Evidence from 20,000 Electoral Slates

This chapter was coauthored with Klára Svitáková (CERGE-EI).

1.1 Introduction

Proportional representation (PR) electoral systems are used in both national and local elections by most EU countries as well as in European Parliament elections. In these systems, political parties have strong gate-keeping power, and thus substantially influence the selection of candidates into mandates. Party leaders selecting candidates to feature on ballots and determining candidates' ranking on party slates are likely to maximize their party's vote share by selecting high valence candidates who presumably appeal to voters. At the same time, party leaders are believed to pursue other within-party goals such as rewarding candidates' loyalty (Galasso and Nannicini, 2017) or defending their own leadership positions within the party (Besley et al., 2017). Political parties can improve selection of candidates into mandates by partially overcoming information asymmetry between candidates and voters (Caillaud and Tirole, 2008), but they also present the voters with a principal-agent problem, in which voters cannot fully control the pre-selection of candidates. In this study, we provide novel evidence and intuition on how these two aspects of slate formation affect the structure and candidates' ranking on party slate. We analyze the ranking of candidates on over 20,000 party slates from five Czech municipal elections, in which the candidates' ranking on the party slate is

informative about the probability of winning a mandate. We categorize candidates in terms of their valence approximated by their education level and by their intra-party value measured by their party membership status and/or party donations. We find that: (a) high valence candidates are placed in better ranked positions than low valence candidates; (b) candidates with high intra-party value are placed in better ranked positions than candidates with low intra-party value; (c) conditional on observables including slate rank and valence, candidates with high intra-party value tend to receive significantly fewer votes than their counterparts with low intra-party value; and (d) an increase in party popularity is associated with a sizeable increase in the share of candidates with high intra-party value and a weak increase in the share of high valence candidates on the party slate.

To explain our findings, we propose a simple model of the market of candidates. A party leader, the demand side of the market, selects and ranks candidates on a slate and values: (i) candidates' valence, as this attracts swing (quality sensitive) voters; and (ii) intra-party value of candidates (provision of scarce resources for the party, e.g. donations or voluntary labor). Potential candidates, the supply side, are either of high or low valence, which entail different opportunity costs of running in the election and they decide on costly actions that can increase their intra-party value (e.g. to become party members or to donate money). As a result, in an environment where the party holds strong gate-keeping power, the party leader trades slate ranks which embody the probabilities of winning mandates in exchange for candidates' valence and intra-party value. Candidates accept the party offer of a slate position if it satisfies their participation constraint. The model yields two main implications. First, candidates who are more valuable to the party are rewarded by better ranked positions. Second, parties that can offer more slate positions with a high probability of winning a mandate, attract more valuable candidates, both in terms of valence and intra-party value.

Our contribution to the existing literature that studies the role of political parties in selection of candidates is threefold.[1] First, we contribute to the recently emerging literature that studies the mechanisms driving candidates' ranking on slates. Literature has emphasized the role of candidates' political experience on their intra-party positions. In particular, Cirone et al. (2020) propose that candidates' intra-party positions, and consequently slate positions, are driven by two rules: *incumbent re-nomination norm*

[1]Dal Bó and Finan (2018) provide a comprehensive summary of recent progress in the literature of political selection.

and *seniority progression norm*. In a similar vein, Fiva and Røhr (2018) show that in party-list systems, the incumbency advantage of candidates is driven by better slate positions, which effectively highlights the importance of political experience for intra-party position and slate position. Studying the role of quality (valence) of candidates on their slate positions, Buisseret et al. (2019) provide robust evidence that in the PR system (similar to the one studied in this chapter), candidates are ranked according to their quality in descending order. We add to this literature by providing novel evidence on how intra-party value, measured by party membership and/or donation, interact with valence of candidates and how it shapes party slate.

Second, we explicitly consider candidates' participation constraints and thus effectively add the supply side of the candidates' market to the framework. While candidates' participation constraint is standard in models of political selection with a focus on the self-selection decisions of candidates, models studying the role of parties in the selection of candidates have neglected it. The candidates' participation constraint allows us to explain an increase in the shares of high valence candidates and candidates with high intra-party value, as the party expects more mandates to win. Third, contrary to the previous literature that assigns candidates one of two mutually exclusive characteristics, i.e. candidates can be either of high valence (experts) or loyal, we relax this assumption and assign candidates two characteristics: levels of valence and of intra-party value. Specifically, we treat the intra-party value as a choice variable of the candidates' problem. Assigning candidates both characteristics allows us to address an apparent controversy in the previous literature. In a study by Galasso and Nannicini (2015), a party leader ranks two mutually exclusive types of candidates on a slate: loyal and expert candidates. The authors show that better ranked positions (safe positions) tend to be occupied by loyal candidates (party officials and incumbent members of parliament). Conversely, Buisseret et al. (2019), using Swedish administrative data, show that candidates are ranked in descending order according to their quality. Since we attribute both characteristics to each candidate: (i) quality (expertise or valence); and (ii) loyalty (intra-party value), we can reconcile both observations through the possibility of ranking high valence candidates with high intra-party value in the top slate positions.

More broadly, this chapter builds on the literature that places political parties and their interests at the center stage of the candidate selection process. Researchers have proposed different reasons why political parties may not strictly prefer high valence candidates. In Besley et al. (2017), a party leader balances the potential threat of being

overthrown by high quality party members against voters' preference for competent candidates. Mattozzi and Merlo (2015) present a model in which having a strong candidate may discourage other candidates from joining the party; therefore, it may be optimal to recruit only mediocre candidates. Alternatively, Galasso and Nannicini (2011) and Galasso and Nannicini (2017) propose that leaders may prefer loyal candidates who, in their models, cannot be of high valence. None of the models allows the authors to address the findings that we document.

The Czech Republic is a convenient case study due to the availability of data, large number of municipalities, the legal option to make political donations and duty to declare them, and the presence of the PR system in which independent candidates (non-members) are allowed to run on party slates.[2] However, we believe the implications of our results are applicable to many national elections and to European Parliament elections.[3]

1.2 Institutional Background and Data

In the Czech Republic, public administration is organized into three levels: central, regional, and municipal. There are more than 6,000 municipalities, and each has its own council and representatives who are elected every four years in municipal elections. The number of mandates in a municipal council depends on the number of citizens in the municipality and varies from 5 in the smallest municipalities to 70 in the capital city of Prague. The number of residents in municipalities varies, averaging around 1,600. Municipalities are responsible for delivering public goods including schooling, municipal infrastructure, and waste management. Czech municipal elections are characterized by large numbers of candidates and parties. When municipal elections are held, there are around 200,000 candidates nationwide running for local mandates. Roughly one third of them will win a council mandate. Generally, about half of the candidates run on the slate of a local branch of a national party, while the rest run on a slate of one of the purely local parties.[4] Local branches of national parties, the focus of this study, are more

[2]Several scholars have used the advantage of the large number of Czech municipalities and publicly available data about: (i) political donation; (ii) all candidates running in municipal and/or regional elections in the Czech Republic. See, for example, Jurajda and Münich (2015); Palguta and Pertold (2018); Palguta (2015); Titl and Geys (2019).

[3]The power of political parties over electoral results in European Parliament elections varies from country to country.

[4]The exact shares of candidates running on the slates of national parties varies by election and depends on the classification of national parties and election coalitions of parties.

professionally organized, whereas local parties, the majority of which are active only in one municipality, often lack effective structural internal organization.

Municipal elections in the Czech Republic are classified as open list elections, which means that parties rank candidates on the slates but voters are allowed to cast votes for their desired candidates. Each voter has as many votes as there are mandates to be allocated. Voters can follow one of three voting strategies. First, they can cast all their votes for one party. Second, they can distribute votes preferentially to different candidates regardless of the slate they are listed on. Third, they can combine the two approaches, i.e. some of their votes can be allocated directly to preferred candidates and the remaining votes to a party. In that case, the remaining votes are assigned to the top ranked candidates on the party slate. The top ranked candidates thus mechanically receive more votes. No one can give more than one vote to any candidate. The number of candidates on the slate of a party is limited to, at most, the number of mandates in the municipal council. The allocation of mandates to parties is determined using the D'Hondt method based on all votes the party received, including those allocated to individual candidates as preferential votes.[5] If a candidate receives at least 110% of the votes of the party average per candidate, then he automatically skips to the top of the slate. Over the past five municipal elections, 15% of mandates were assigned to candidates who received enough preferential votes to skip higher in the ranking, and who would not have won the mandates otherwise. 15% is not insignificant, but it is clear that the initial party ranking substantially shapes the final electoral outcome, as the remaining 85% of mandates were assigned to the candidates at the top of the slate, i.e. those pre-selected by the party. In fact, well ranked candidates can be elected even when there are other not-elected candidates in lower positions on the slate who receive more votes, but not enough to skip to the top.

The Czech legal system allows both individuals and firms to make donations to political parties. A complete list of political donors, including additional individual information, is required to be published by the political parties annually. We collect the data on donations made by individuals and firms between 1995[6] and 2018 and match them with a dataset of all candidates in all elections since 2002. This allows us to identify candidates who donated money to the party on whose slate they run and to classify them

[5]Note that there is also a threshold share of all valid votes that the party has to exceed, otherwise it is not given a mandate. The default threshold is 5%, and it can be lower for parties that have fewer candidates than there are mandates to allocate in the municipality.

[6]Prior to 1999, parties did not have to publish donations of less than CZK 100,000.

as candidate-donors.[7]

The available data consist of a universe of individual candidates for each election from 1998 to 2018. We observe each candidate's name, age, academic degree, place of residence, occupation, party membership, the party they run for, position on the slate, the number of votes received, and elected status.[8] To create a panel structure, we match candidates across different types of elections (municipal, regional, parliamentary) and different election years. Unfortunately, the candidates do not have individual unique identifiers, so instead, we match them using their individual characteristics including name, surname, year of birth, education level and, where possible, place of residence. Since the name is one of the main characteristics that we use for matching, it is more complicated to correctly match female candidates, as their surnames may change after marriage. We perform robustness checks by matching female candidates using all the usual characteristics except for surname, and none of the analysis changes. The initial dataset consists of 735,393 unique individuals who have run in at least one election since 1998. We restrict the dataset to candidates who have run in at least one municipal election for one of the six largest national parties (KDUCSL, CSSD, KSCM, ODS, TOP09, ANO)[9] in one of the last 5 municipal elections (2002, 2006, 2010, 2014, and 2018).[10] Additionally, in order to ensure comparability across slates, we drop all candidates who ran on incomplete slates, i.e. slates that list fewer than the maximum possible number of candidates.[11] We end up with a dataset consisting of 214,580 individuals. Some of the candidates are observed repeatedly, as they run in several elections. Table 1.7 in Appendix A summarizes the numbers of candidates running for different national parties.

1.3 Empirical Evidence

1.3.1 Types of Candidates

The order of candidates on the slate is determined by many aspects including the characteristics of the candidates (e.g. political experience and ability), internal party organi-

[7] We link the donations of firms to their owners, executive directors, or board members who run for office.

[8] Occupation and place of residence are self-reported.

[9] Note that TOP09 only participated in the last three elections and ANO in the last two elections.

[10] We do not consider candidates who run on a joint slate for two or more parties in a coalition, as we do not observe which party nominated which candidate.

[11] The maximum possible number of candidates on a slate equals the number of council seats in a municipality.

zation (who bears responsibility for slate formation and their preferences), municipality and voters' characteristics, and political competition. We explore the roles of candidates' valence and their intra-party value, and document that both play a major role in explaining the observed ranking of candidates on slates. Intuitively, valence represents the public value of candidates, i.e. it is the characteristic that voters care about, while intra-party value is any characteristic that the party itself appreciates.

We classify the valence of politicians by their education level. Specifically, we consider candidates as being of high valence if they have obtained at least a college degree and of low valence otherwise. This approach is standard in the literature of political selection (e.g., Dal Bó et al., 2009; Ferraz and Finan, 2009).[12] Importantly, Buisseret et al. (2019) show that education displays similar patterns on slates as other (likely better) measures of the quality of politicians such as perceived leadership ability, cognitive scores and labor market income, providing some support for our use of the measure. Using education as a measure of valence can be problematic, as education combines other characteristics of an individual candidate. Furthermore, since the data does not allow us to take the quality and particular specialization of education obtained into account, it is a noisy signal correlated with valence. Nevertheless, there is little evidence or consensus among the general public and researchers on what characteristics qualify politicians as high valence and even less so when the discussion is restricted to measurable and available characteristics.

We use two distinct measures to quantify the intra-party value of candidates: (i) party membership status; and (ii) party donations. Candidates in any election can run on a party's slate even if they are not formal party members. On the slate, such candidates are labeled "without political affiliation". Candidates who are party members of any political party are labeled with the party name. Being a member of a political party often comes with costs. At the very least, all members usually have to pay a membership fee. Further, they can take on other duties and work for the party, and may provide voluntary labor and help with fundraising, organization, and campaign activities. A candidate is classified as a party donor if: (i) he or a firm that he owns or represents is listed as a donor by the party he runs for; and (ii) the timing of the donation is close to the election, specifically the year prior to municipal elections, the election year, or one year after.[13]

[12] Dal Bó et al. (2017) argue that while education is correlated with ability, it may also reflect luck or social class.

[13] The results are robust to different specifications of the time window.

Table 1.1 presents shares of different types of candidates in our dataset. Roughly 27% of candidates are classified as high valence candidates and 73% as low valence candidates. More than 40% of candidates are party members, while only 2.5% are party donors. The significant difference can be interpreted as a consequence of donations being a more costly form of intra-party value for candidates compared to active membership status. We later discuss the difference in more detail.

Table 1.1: Shares of Types of Candidates

	High Valence			Low Valence		
	Member	Non-Member	Total	Member	Non-Member	Total
Donors	1.40%	0.21%	1.61%	0.93%	0.15%	1.08%
Non-Donors	10.93%	14.79%	25.73%	29.51%	42.07%	71.58%
Total	12.33%	15.01%	27.34%	30.44%	42.22%	72.66%

1.3.2 Slate Structure

The number of candidates on slates differs across municipalities, parties, and election years. In order to compare the ranking of candidates across different slates, we define *Rank* as the position on the slate: (i) conditional on other observable characteristics, e.g. political experience, age, and nominating party; and (ii) normalized to be within the [0,1] interval, where 0 is the top position on the slate and 1 is the bottom. The former helps us to concentrate on the two dimensions of candidates' characteristics that we study. As part of the robustness exercises, Appendix A replicates the figures in this section using the unconditional raw rank and confirms the same patterns. We use this conditional normalized measure of *Rank* throughout this section.[14]

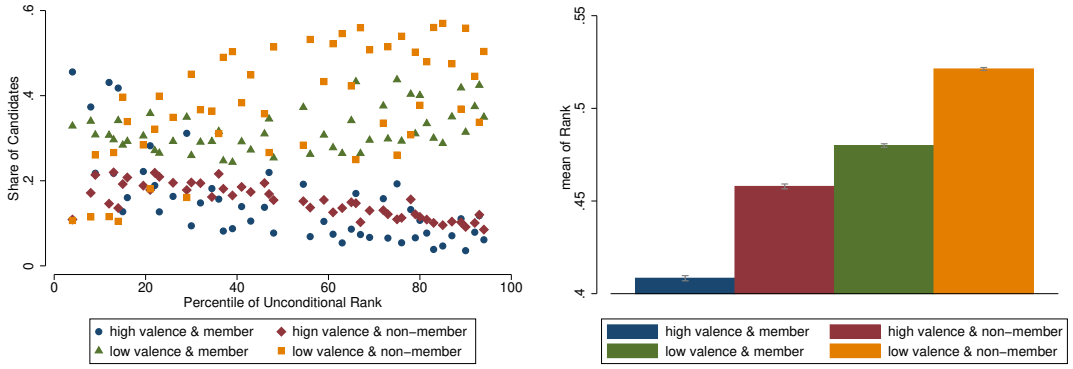
Party Members

We first document that candidates classified according to their valence and party membership status are systematically ranked on the slate. High valence candidates and party members are over-represented in better ranked positions, i.e. positions with a higher probability of being elected, and are under-represented in worse ranked positions. Observation 1 summarizes the pattern in terms of the average *rank* of different groups.

[14]For more details about *Rank*, see Appendix A. Appendix A also provides several exercises to demonstrate that our results are robust to different specifications.

Observation 1 (Slate Structure - Party Members)**.** *Party members are systematically ranked on slates. In terms of average rank, the groups are ranked as follows: (i) high valence members at the top; followed by (ii) high valence non-members; (iii) low valence members; and (iv) low valence non-members at the bottom of the slate.*

Figure 1.1: Slate Structure for Members



(a) Slate Structure (Members) **(b)** Average Rank by Types (Members)

Figure 1.1a graphically represents shares of different candidate types on different slate ranks. The x-axis shows a percentile of variable *Rank*, while the y-axis shows the shares of the different groups of candidates within the corresponding percentile of *Rank*. Roughly a third of the candidates placed in the top 5 percentiles of the *Rank* are classified as high valence party members; another third consists of low valence party members; and the rest is occupied by non-members of both levels of valence. Towards the bottom of the slate, the share of low valence non-members increases, while the share of high valence candidates, both members and non-members, decreases. Overall, as we move from the top ranked slate positions to the bottom of the slate, high valence candidates (both members and non-members) are gradually replaced by low valence candidates (both members and non-members). The same is apparent for party members (both high and low valence), who are over-represented among the better ranked positions. Figure 1.1b summarizes the average of variable *Rank* and confidence intervals of the four types of candidates and confirms Observation 1. Appendix A presents two robustness exercises that confirm the same ranking pattern among candidates with no previous political experience and for candidates running on specific slates that list at least one candidate of each type.

Interestingly, the bottom of the slate shows a drop in low valence non-members and peaks in all three remaining groups. There are two possible explanations. First, it may be

that some popular politicians from the national parliament, local celebrities, or respected residents with no interest in being elected in municipal elections are voluntarily placed at the bottom in order to attract voters' attention to the party. When they are elected, they often refuse the mandate, as their main motivation for running is not to be elected, but rather to support the party. Second, voters may pay more attention to the candidates at the bottom of the slate than to those around the middle of the list. Some candidates may consider the bottom position more visible and thus more likely to attract preferential votes.

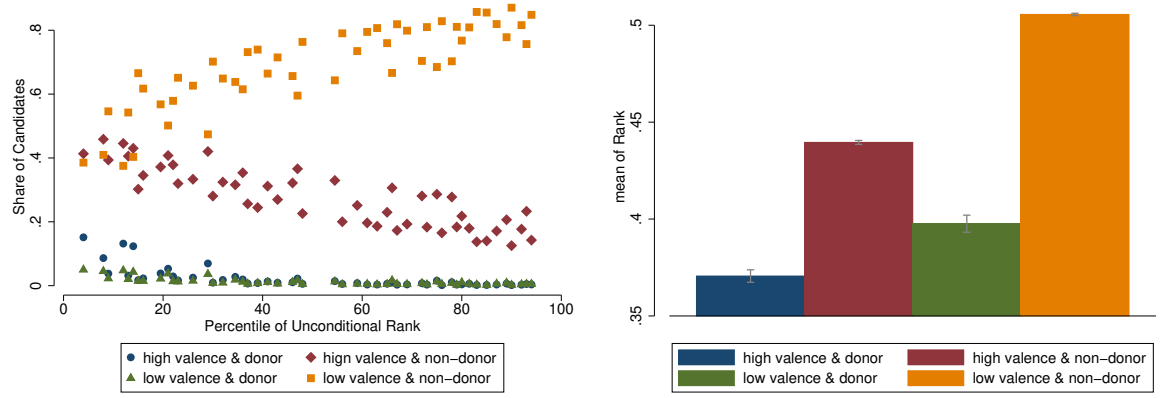
Party Donors

Instead of party membership status, we next use party donations as a measure of the intra-party value of candidates. This leads to a new classification of the four groups: high valence donors; high valence non-donors; low valence donors; low valence non-donors. The main ranking pattern using this new classification resembles the pattern for party members; donors are, on average, ranked better than non-donors, as is summarized in Observation 2.

Observation 2 (Slate Structure - Party Donors)**.** *Party donors are systematically ranked on slates. In terms of average rank, the groups are ranked as follows: (i) high valence donors; (ii) low valence donors; (iii) high valence non-donors; (iv) low valence non-donors.*

The pattern persists with one notable exception. As expected, high valence party donors are over-represented in the best ranked positions and under-represented in the worst ranked positions, while the opposite is true for low valence non-donors. However, the two middle groups switch their positions; low valence donors are ranked better than high valence non-donors. Applying an alternative, and arguably more costly, measure of intra-party value leads to a switch between the two types of candidates: low valence candidates with more intra-party value tend to be in better positions than high valence candidates with low intra-party value. Figure 1.2b provides a graphical representation of Observation 2. Similarly to Figure 1.1a, Figure 1.2a shows a spike in high valence candidates and party donors at the bottom of the slate.

Figure 1.2: Slate Structure for Donors



(a) Slate Structure (Donors)

(b) Average Rank by Types (Donors)

1.3.3 Intra-party Value and Slate Rank

We next provide additional evidence that candidates with high intra-party value are ranked better on the party slate. In particular, both party membership status and party donations are associated with better ranked positions and higher probability of being placed in an electable position. Observation 3 summarizes the findings.

Observation 3 (Intra-party Value of Candidates)**.** *Becoming a member and/or a donor is associated with a shift towards better ranked positions and an increase in the probability of being placed in an electable position.*

The data are organized in an unbalanced panel with an individual candidate in a given election year being the unit of observation. If a candidate does not run in a certain year, he is missing from the data in that year. On average, candidates run in 1.6 elections. The outcome variables of our interest are: (i) slate position; and (ii) indicator for being placed in an electable position. The slate position of an individual candidate is normalized to be between 0 and 1[15], such that the first position on the slate is always ranked 0, and the last position ranks 1. The indicator for electable position equals 1 if the candidate's slate rank would win a seat if the party received as many seats as it did in the previous elections, and 0 otherwise.

We report results from pooled OLS and from a fixed effects model regressions. In both, we control for candidate characteristics such as age and previous political experience in municipal, regional, parliamentary, and senate elections[16], and for incumbency status,

[15] We use the transformation Rank = (Slate position-1)/(Total number of candidates-1).

[16] Both are controlled for using fixed effects.

using an indicator equal to 1 if the candidate won a seat in the previous elections. We assign a particular donation to given elections if it was made the year prior to the elections, the year of the elections, or the year after. We remove candidates who simultaneously run for other offices during the political cycle and whose donation windows overlap, because their donation could be related to different elections. We also include party and time fixed effects and their interaction. See equation 1.1, where δ_{it}^{q} stands for fixed effects of the number of previously held mandates; ω_{it}^{q} stands for fixed effects of the number of previous candidacy; γ_i is an individual fixed effect; and depending on specification Y_{it} stands either for an outcome variable $Normalized\ Rank$ or an indicator whether the candidate was placed in electable positions $Electable\ Position$.

$$\begin{aligned} Y_{it} = \beta_1 Donation_{it} + \beta_2 Donation\ Size_{it} + \beta_3 Membership_{it} + \beta_4 Gender_{it} + \\ + \beta_5 Degree_{it} + \beta_6 Age_{it} + \beta_7 AgeSq_{it} + \delta_{it}^{q} + \omega_{it}^{q} + \delta_{it}^{q} \times \omega_{it}^{q} + \gamma_i + \varepsilon_{it} \end{aligned} \tag{1.1}$$

Column 1 in Table 1.2 reports estimates from the pooled OLS regression with $NormalizedRank$ as the dependent variable and shows that the coefficient on the Donation Dummy variable is negative, suggesting that, on average, donors are placed in better ranked positions. Surprisingly, the sign of the coefficient of the Donation Size variable is positive and works in the opposite direction than the dummy indicator. For the overall effect of a donation to be positive, one would have to make a party donation of CZK 3.5 million, which is rather rare in our data.[17] The effect of party membership status is of the expected sign but of lower magnitude. As expected, academic degree (our measure for valence of candidates) and incumbency status are also associated with a better ranked position. Column 2 shows how individual characteristics of candidates are related to the probability of being placed in electable positions. Party donation is associated with a 12 percentage point increase in the probability of being placed in an electable position and party membership status appears to be linked negatively to the probability of being placed in electable positions. The counter-intuitive effects of the intensive margin of party donation on rank and party membership status on the probability of being placed in electable positions are likely driven by unobserved characteristics of candidates.

We next report results from the fixed effects model that controls for the time invariant unobservable characteristics of candidates such as motivation, ability, and local

[17]There are only 14 donations of CZK 3.5 million or higher in our data.

Table 1.2: Intra-party value Improves Slate Rank

	(1) Normalized Rank	(2) Electable Position	(3) Normalized Rank	(4) Electable Position
Party Donation Dummy	-0.130***	0.122***	-0.053***	0.099***
	(0.003)	(0.005)	(0.004)	(0.007)
Donation Size (in millions CZK)	0.037***	0.012	-0.002	0.044**
	(0.010)	(0.013)	(0.006)	(0.017)
Party Membership	-0.072***	-0.007***	-0.105***	0.073***
	(0.001)	(0.001)	(0.003)	(0.005)
Academic Degree	-0.116***	0.018***		
	(0.001)	(0.002)		
Incumbent	-0.170***	0.273***	-0.134***	0.184***
	(0.003)	(0.005)	(0.004)	(0.007)
Individual FE	No	No	Yes	Yes
Age	Yes	Yes	Yes	Yes
Political experience FE	Yes	Yes	Yes	Yes
Party and year FE	Yes	Yes	Yes	Yes
N	345,701	345,701	345,701	345,701

Standard errors in parentheses

Party and year fixed effects include their interactions.

Previous political experience includes running and receiving a mandate in municipal elections, regional elections, parliamentary elections and/or senate elections.

* $p < .10$, ** $p < .05$, *** $p < .01$

popularity. Results presented in columns 3 and 4 in Table 1.2 are more in line with our expectations, as both the party donation dummy and party membership status have the expected signs. Becoming a party member and a party donor is associated with better ranked positions on the slate and with higher probability of being placed in electable positions. For example, donating CZK one million is associated with a 14.3 percentage point higher likelihood of being placed in an electable position, and becoming a member increases the likelihood by 7.3 percentage points. To interpret the size of the effect, consider a slate of 15 candidates, which represents a slate of median length. Becoming a party donor leads to a rank improvement by almost one position and becoming a party member by a little over 1.5 positions. We cannot rule out that the results are driven by some time varying characteristics, such as an increased interest in a political career, that would place the candidate in better positions on the slate and at the same time increase his likelihood of becoming a party member and/or party donor.

1.3.4 Intra-Party Value and Electoral Performance

We next provide evidence that, conditional on the level of valence, slate position, and other observable characteristics, candidates with higher intra-party value receive fewer votes than their counterparts with lower intra-party value. Candidates who are valued by party leaders for their intra-party value appear not to be equally popular among voters. Our results thus suggest that party leaders face a trade-off between intra-party value of candidates and candidates' attractiveness among voters. We argue that the effect is driven by negative selection on characteristics that are unobservable to us, but observable to voters at the time of election. Observation 4 summarizes our finding.

Observation 4 (Electoral Performance of Candidates)**.** *Conditional on slate position and other characteristics of candidates and the party, candidates with higher intra party value tend to receive fewer votes.*

Since the electoral system mechanically favours better ranked candidates, a simple comparison of votes cast for different candidates on the same slate is not informative about candidates' electoral performance. Instead, we define a candidate's $Relative\,Votes$ and compare this measure among candidates in the same slate position but on different slates. The measure $Relative\,Votes$ is defined as a ratio of votes a candidate i received and the slate's average number of votes per candidate (a candidate who receives the

average number of votes has $Relative\,Votes_i$=1). [18]

The fact that slates differ in their length complicates the comparison even further. We approach this problem by: (i) using a flexible function of the normalized rank, using observations from all the slates available regardless of their length; (ii) studying only slates with a specific length (with 15 and 21 candidates) and using dummy variables for each slate position. Table 1.3 shows results from all three specifications. In all specifications, we control for candidates' previous political experience, gender, and age. See equation 1.2.

$$Relative\,Votes_{it} = \alpha + \beta_1 Membership_{it} + \beta_2 Donation_{it} + \beta_3 Degree_{it} + \\ + \beta_4 Slate\ Position_{it} + \beta_5 Gender_{it} + \beta_6 Age_{it} + \beta_7 AgeSq_{it} + \delta^q_{it} + \omega^q_{it} + \delta^q_{it} \times \omega^q_{it} + \varepsilon_{it} \tag{1.2}$$

Table 1.3: Intra-party Value and Electoral Performance

	(1) Relative Votes	(2) Relative Votes	(3) Relative Votes
Party Membership	-0.045***	-0.058***	-0.064***
	(0.001)	(0.002)	(0.002)
Party Donation Dummy	-0.061***	-0.034***	-0.033***
	(0.003)	(0.005)	(0.007)
Academic Degree	0.076***	0.066***	0.085***
	(0.001)	(0.002)	(0.002)
Normalized Rank	Yes	No	No
Slate Positions Dummies	No	Yes	Yes
Types of Slates	All	21 candidates	15 candidates
N	343,991	61,575	120,068

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

The first column of Table 1.3 shows results from a specification in which we use slates of all lengths and control for a polynomial function of normalized rank ($Normalized\,Rank$) of candidates. The coefficients suggest that party members receive 4.5 percentage points fewer $Relative\,Votes$ than a party non-member would receive keeping everything else

[18] Figure 1.5 in the Appendix shows average $Relative\,Votes$ for slates with 15 candidates.

the same. Suppose, for example, that the predicted *Relative Votes* for a top ranked position are 140% of the slate average votes. Then, if a party member runs on that slate position, he receives only 135.5% of the slate average votes. Similarly, party donors receive 6.1 percentage points fewer *Relative Votes* than non-donors. The second and third columns show the results for slates with 21 and 15 candidates, respectively. In these specifications, we control for a vector of dummy variables corresponding to each slate position. The results resemble those from the first column; party members and party donors receive significantly fewer *Relative Votes* than their counterparts.

There are two possible explanations as to why members and donors received relatively fewer votes conditional on other observed characteristics. First, voters dislike members and donors, and second, candidates with intra-party value are negatively selected based on some characteristics that are unobservable to us, but observable to voters at the time of the elections. While we cannot rule out any of the explanations, we consider the latter much more plausible. The main argument against the former explanation is that the list of political donors is publicly available only a year after the election. Donors are thus rarely known at the time of the election. We thus consider the latter explanation more credible. Donors and members tend to be negatively selected and differ in some, for us unobserved, characteristics such as individual quality, reputation, political scandals, charisma, and credibility that are, however, observable to voters. Similarly, donors and members may be less motivated and exhibit less effort during the electoral campaign. Regardless of the channel through which the negative selection of donors and members operates, from the party leader's perspective, it is important that candidates with intra-party value under-perform and receive fewer votes than their counterparts.[19]

1.3.5 Popular Parties and Valuable Candidates

We next show that a party's popularity affects the slate structure. Specifically, as a party becomes more popular and expects more mandates, there are more valuable candidates on the slate. Suppose there is a popularity index for each party at the municipal level. This variable is, at least to some extent, visible to the voters, but it remains latent to us. Our only observable realization is through the shares of votes during elections. As

[19]Nevertheless, it is still possible and consistent with our results that placing a donor on a party slate could lead to a better electoral outcome than the party would reach without the donor. For example, if the funds donated enable the party to run a campaign that attracts a mass of voters and the funds would not be available without the donor being placed on the party slate, it may be worth keeping the under-performing donor on the slate.

the popularity of the party increases, so does its share of votes. We measure a party's popularity by the share of votes the party received in the most recent parliamentary election at the municipal level. We show that, after a party becomes more popular, which implies that it can expect to win more mandates, it weakly places more high valence candidates on the slate and place significantly more candidates with high intra-party value on the slate. Observation 5 is thus consistent with the interpretation that a more powerful party can attract more high valence candidates and prompt them to increase their intra-party value (i.e. become a member and/or a party donor).

Observation 5 (Popularity of Parties)**.** *After a local popularity shock, there are weakly more high valence candidates and significantly more candidates with high intra-party value on the slate. In particular, the share of high valence candidates with high intra-party value increases, while the share of low valence candidates with low intra-party value decreases. This is the case for both of our measures: party membership status and party donations.*

To measure the popularity of parties and changes in the party's popularity at the municipal level, we rely on parliamentary election results. This is possible because data about party vote shares in the parliamentary elections are available at the municipal level and because parliamentary elections conveniently take place from 4 to 12 months prior to municipal elections. Figure 1.3 shows the sequence of parliamentary and municipal elections in different years. Our specification (Equation 1.3) controls for time-party and municipality-party fixed effects, and the identification is thus based on the time variation in municipal political preferences that is orthogonal to changes in national political preferences and to long-term geographical variation in political preferences. For example, the local perception of national or regional policies promoted by a given political party generates such variation.[20] Furthermore, we control for time-varying slate structures at the regional level, and thus any within-party organizational changes (e.g. party level demand for donors) in slate formation are filtered out.

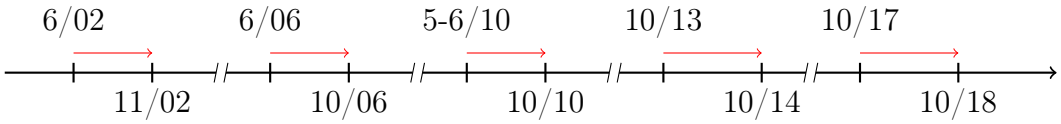
For both of our measures of intra-party value we run the following regression.

$$Share^{g}_{pj\tau} = \alpha^{g} + \beta^{g} PE\ ShareVotes_{pj\tau} + \sum_{k \in \{HM,HN,LM\}} \delta^{k} PE\ Share^{k}_{p\tilde{j}\tau} + \gamma^{g}_{pj} + \gamma^{g}_{p\tau} + \epsilon^{g}_{pj\tau} \tag{1.3}$$

[20]National policies promoted by a given political party may affect municipalities differently depending on their local demographic and economic conditions.

Figure 1.3: Sequence of Elections

Parliamentary Elections:



Municipal Elections:

where p denotes political party, j municipality, τ is a political cycle, i.e. a sequence of parliamentary and municipal elections, and k is a type of candidate: high valence with high intra-party value (HM,HD), low valence with high intra-party value (LM,LD), high valence with low intra-party value (HN), and low valence with low intra-party value (LN). $PE\ ShareVotes_{pj\tau}$ is the share of votes that a party p received in municipality j in the parliamentary elections during a political cycle τ, and finally $PE\ Share^{k}_{p\tilde{j}\tau}$ captures the share of candidates of group k on the slate of party p in the parliamentary elections in the electoral region $\tilde{j}$ and political cycle τ. In parliamentary elections, parties form an individual slate in each of fourteen regions $\tilde{j}$, and each municipality j belongs to one region. We include $PE\ Share^{k}_{p\tilde{j}\tau}$ to control for the effect of the structure of the slate in the particular region, i.e. to control for the possibility that a party receives more votes in a given municipality not because it gained popularity, but because it formed a particularly good slate in the parliamentary elections.

Party Membership

We first discuss results for party membership status as a measure of the intra-party value of candidates. An increase in a party's share of votes in a parliamentary election is associated with an increase in the number of party members on the slate in the subsequent municipal election. Formally, we run Regression 1.3 separately for each of the following types of candidates g: (i) high valence members (HM); (ii) high valence non-members (HN); (iii) low valence members (LM); and (iv) low valence non-members (LN).

Each column of Table 1.4 represents a regression for one group of candidates. The first row captures estimates of β^{g} from Equation 1.3. A one percentage point increase in the vote share in a parliamentary election in a given municipality is associated with an

Table 1.4: Changes in Party Popularity and Shares of Members

	(1) Share of HM	(2) Share of HN	(3) Share of LM	(4) Share of LN
PE Share Votes	0.080***	-0.033	0.352***	-0.400***
	(0.017)	(0.024)	(0.033)	(0.036)
N	21,442	21,442	21,442	21,442
Party Year FE	Yes	Yes	Yes	Yes
Party Municipality FE	Yes	Yes	Yes	Yes
PE Share of HM, HN, and LM	Yes	Yes	Yes	Yes

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

increase of 0.08 percentage points of the share of high valence members in the subsequent municipal election. Since the average share of high valence members is roughly 10 percent of a slate, the effect represents a 0.8% increase. The results further show that the share of low valence members increases by 0.35 percentage points and the share of low valence non-members decreases by 0.40 percentage points. These effects represent a 1.2% increase in low valence members and a 0.85% decrease in low valence non-members, respectively. Overall, low valence non-members, who are arguably the least valuable to the party leader, are squeezed out and replaced by more valuable types of candidates as the party's popularity increases. An increase in the vote share in a parliamentary election is followed by a municipal election slate that includes more high valence candidates and strictly more party members. Considering a slate of a median length, i.e. 15 candidates, receiving an additional 14 percentage points of votes in parliamentary elections implies one additional member in the subsequent municipal election.

Party Donors

The effects for party donors are qualitatively equivalent but of a lower magnitude. An increase in the vote share of a party in a parliamentary election is connected to an increase in the shares of high and low valence donors, while the share of the least valuable candidates, low valence non-donors, decreases. This implies an increase in both the share of donors and the share of high valence candidates. Formally, we run regression (1.3) for g: (i) high valence donors (HD); (ii) high valence non-donors (HN); (iii) low valence donors (LD); and (iv) low valence non-donors (LN). Table 1.5 shows the results.

Table 1.5: Changes in Party Popularity and Shares of Donors

	(1) Share of HD	(2) Share of HN	(3) Share of LD	(4) Share of LN
PE Share Votes	0.019**	0.029	0.028***	-0.076***
	(0.009)	(0.027)	(0.009)	(0.027)
N	21,442	21,442	21,442	21,442
Party Year FE	Yes	Yes	Yes	Yes
Party Municipality FE	Yes	Yes	Yes	Yes
PE Share of HD, HN, LD	Yes	Yes	Yes	Yes

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Receiving 10 additional percentage points in a parliamentary election is related to a 0.19 percentage point increase in high valence donors on the slates. Taking the average slate structure as a baseline case, i.e. including only 1.4% of high valence donors, this is equivalent to a 1.3% increase in the number of high valence donors. The most pronounced positive effect is among low valence donors, as an increase of 10 percentage points in parliamentary elections implies 0.28 percentage points, or a 2.8% increase in the share of low valence donors on average. The increase in the share of donors is offset by the share of low valence non-donors, whose share falls by 0.76 percentage points after a 10 percentage point popularity shock. In Appendix A, we show that the effects are similar for all political parties studied and robust to the different source of variation in a party's popularity.

1.4 Interpreting the Results

1.4.1 Theoretical Framework

In this section we build a highly stylized model of the candidate selection process and use it to formalize the intuition for the ranking patterns observed on slates. We think of the selection process as a market of candidates in which a party leader (she) demands candidates' valence and intra-party value in exchange for slate positions, while candidates offer their valence and costly intra-party value in exchange for the probability of winning a mandate. The party leader forms the slate and decides what types of candidates will be placed at what positions on the slate. Her objective is twofold. First, to attract swing voters and thus increase the chances of success in elections, she needs high valence

candidates on the slate. Second, as for her intra-party objective, she maximizes the number of candidates with high intra-party value. We will consider the problem of a single political party and omit interactions between different parties.

For convenience, we normalize the slate length to an interval [0,1] and denote a slate rank as $t \in [0,1]$, such that $t = 0$ is the top rank and $t = 1$ the bottom. Any candidate placed on a slate rank t has two indicator characteristics: (i) valence v; and (ii) intra-party value m. If a candidate placed on t rank is of high valence, then $v(t) = 1$, otherwise $v(t) = 0$. Similarly, if a candidate placed on t rank has high intra-party value, then $m(t) = 1$, otherwise $m(t) = 0$. The key object of our framework is a slate characterized by $(v(t), m(t))$, where $v(t) : [0,1] \mapsto \{0,1\}$ and $m(t) : [0,1] \mapsto \{0,1\}$, so it maps each slate rank t into a space of the characteristics of the candidates placed in that position.

Voters For tractability, we highly simplify voters' behavior. As is common in the literature, we assume there are two types of voters: (i) party core voters; and (ii) swing voters. Core voters always vote for their preferred party and thus the party receives α votes from its core voters. The decisions of swing voters depend on the overall valence of the slate. We assume that voters are more sensitive to the valence of the top ranked candidates than to that of those at the bottom of the slate.[21] Specifically, swing voters care about an aggregate measure (weighted average) of the valence of the slate $\bar{v} = \int_0^1 g(t)v(t)dt$, where $g(t)$ is a weighting function satisfying $g'(t) < 0$ and $g(1) > 0$. The party receives $\delta\bar{v}+\epsilon$ votes from swing voters, where ϵ is random noise with a mean of zero. The behaviour of voters therefore yields the following probability of winning a mandate.

$$P(\text{winning a mandate}|\alpha, t, \bar{v}) = P(\alpha + \delta\bar{v} + \epsilon \geq \omega_t) \tag{1.4}$$

where ω_t is a unique threshold for a rank t. The probability is increasing in α and $\bar{v}$, but decreasing in t, as ω_t is increasing in t. Any model of voting behavior with these characteristics is consistent with our framework. Importantly, the individual candidate's probability of winning a mandate is a function of the party's popularity (α), the candidate's slate rank (t), and the overall aggregate valence of the slate ($\bar{v}$). A crucial aspect

[21] There are two reasons to support this assumption. First, even under an open-list electoral system, top ranked candidates are more likely to be elected due to mechanical reasons, as mandates are allocated from the top down. Hence, being more sensitive to the top ranked candidates follows from maximizing the expected valence of elected candidates. Second, if voters are inattentive, they are likely to pay attention to the more pronounced or salient candidates, i.e. the candidates at the top of the list. This assumption has been empirically supported by Buisseret et al. (2019) using Swedish data. They argue that slates are formed according to the *rank-order hierarchy*.

of our setup is, therefore, that voters do not care about intra-party value, only about valence.[22]

Candidates There are two infinitely large pools of candidates: high valence candidates (with $v = 1$) and low valence candidates ($v = 0$), who differ in their opportunity cost of running; $c^h > c^l = 0$, so that candidacy is more costly for high valence candidates. We set the cost of running for low valence candidates at zero.[23] In order to ensure a better slate position, candidates can perform a costly action a, pay cost c^a and become intra-party valuable ($m = 1$). This can take the form of an active party membership status ($a = M$) or a financial donation to the party ($a = D$). Candidates value a mandate that brings them a benefit b, and they maximize their expected payoff (expected benefit minus cost).

Party leader Party leader forms a slate and seeks to maximize her value function

$$V(\bar{v}, \bar{m}) = \bar{v} + \gamma^a \bar{m},$$

where $\bar{v} = \int_0^1 g(t)v(t)dt$ is the measure of overall valence of the slate that follows from the electoral success motive. As $\bar{v}$ increases, so does the expected number of mandates. Additionally, the party value function is increasing in the share of candidates on the slate with high intra-party value, $\bar{m} = \int_0^1 m(t)dt$. The coefficient γ^a captures the relative importance of $\bar{m}$ compared to $\bar{v}$, and may depend on the particular form of intra-party value applied. A crucial property of the party leader's objective is that it is strictly increasing with every additional high valence candidate and with every additional candidate with high intra-party value, holding the rest of the slate constant.[24]

At time $s = 1$, candidates receive an offer from the party leader to run in a particular position on the slate conditional on having a certain intra-party value, and they must decide whether to accept or reject the offer. When making the decision, candidates compare the expected payoff $P(\alpha, \bar{z}, t)b$ with the cost of running and, if required, the

[22] In our setup, this assumption is easier to justify in the case of political donations, which are not visible to voters at the time of the elections, but we find no reason for it to not be true for membership.

[23] This ensures that some candidates are willing to run even in the bottom positions with zero probability of being elected. $c^l > 0$ could lead to incomplete slates.

[24] We assume a very simple value function which is additively separable in valence and intra-party value and where the value of each candidate with high intra-party value is constant. We could also assume that the value of each candidate with high intra-party value is decreasing in his rank, t, as we do for valence. If instead of $\gamma\bar{m}$ we had $\tilde{m} = \int_0^1 \tilde{\gamma}(t)m(t)dt$, where $\tilde{\gamma}'(t) < 0$ and $\tilde{\gamma}(1) \geq 0$, as long as $g(0) > \tilde{\gamma}(0)$ and $\tilde{\gamma}'(t) > g'(t)$ for $\forall t$, the results would be qualitatively unchanged.

cost of becoming highly intra-party valuable. For the party leader, the offer is binding, and she cannot change it once it is accepted by a candidate. Importantly, at the time of the decision, candidates do not know the realized valence $\bar{v}$ of the slate. Instead, they base their decisions on an exogenous prior belief $\bar{z}$. We impose the exogeneity of the candidates' beliefs in order to keep the model as tractable as possible. At time $s = 2$, the party leader assigns positions to candidates given their valence and their affiliation status, and the aggregate valence of the slate $\bar{v}$ is revealed. At time $s = 3$, the election takes place, votes are realized and mandates are assigned to elected candidates.

1.4.2 Characterization of the Solution

There are four thresholds that fully characterize the optimal slate. Three of the thresholds (t_1, t_2, and t_3) represent the supply side of the market and are defined by the participation constraints of candidates, defined by Equations (1.5) - (1.7).

$$P(\alpha, \bar{z}, t_1)b = c^h + c^a \tag{1.5}$$

$$P(\alpha, \bar{z}, t_2)b = c^a \tag{1.6}$$

$$P(\alpha, \bar{z}, t_3)b = c^h \tag{1.7}$$

Each threshold represents the worst slate rank for which the corresponding type of candidates is willing to run. For example, for a high valence candidate with low intra-party value (Equation 1.7) the cost of running is c^h; the worst position that ensures that the expected benefit will be at least equal to the cost of running is rank t_3. As a result, this candidate accepts an offer of slate rank t_3 or lower (i.e., better position). Similarly, the threshold for high valence candidates with high intra-party value is t_1 (Equation 1.5) and for low valence candidates with high intra-party value t_2 (Equation 1.6).

The fourth condition follows from the party leader's preferences and represents the demand side of the market of candidates. Her objective function implies two dominant strategies: (i) she always prefers high valence candidates with high intra-party value over anyone else; and (ii) she always prefers anyone else over low valence candidates with low intra-party value. The only trade-off occurs between high valence candidates with low intra-party value and low valence candidates with high intra-party value in the domain of the slate where both types are willing to run. Holding the rest of the slate constant, the marginal value of the valence of a candidate is $g(t)$. Since voters are more sensitive to the

valence of the top ranked candidates, $g(t)$ is decreasing in the slate rank. On the other hand, the marginal value of high intra-party value of a candidate is γ^a, which is constant across all slate ranks, *ceteris paribus*. Therefore, there is a unique rank, which we denote t_4, for which the party leader is indifferent between high valence candidates with low intra-party value and low valence candidates with high intra-party value. Formally,

$$g(t_4) = \gamma^a. \tag{1.8}$$

For all slate ranks lower than t_4 the party leader prefers high valence candidates with low intra-party value, while for all higher slate ranks she prefers low valence candidates with high intra-party value.

The thresholds might not fall within the [0,1] interval and in that case one or more of the candidate types will not be on the slate at all. We are interested in the general case in which all types are present, so we assume the interior solution (all thresholds are within the [0,1] interval). In appendix B we prove that t_4 as defined in Equation 1.8 maximizes the party leader's value function.

1.4.3 Explaining the Observations

We next link the predictions of our theoretical framework with the established empirical observations. The framework introduced here predicts that the ranking of candidates depends on how the thresholds are ordered. The observed ranking of party members is summarized in Observation 1 which states that, on average, high valence members (HM) tend to be placed at the top of the slate, followed by high valence non-members (HN), low valence members (LM) and lastly low valence non-members (LN). Proposition 1 introduces an equivalence relation between threshold ordering and the patterns observed for party members.

Proposition 1 (Membership)**.** *Consider membership as a measure of intra-party value. If and only if* $t_1^M < t_3^M < t_2^M$ & $t_1^M < t_4^M$*, the group ordering is as follows: (i) HM; (ii) HN; (iii) LM; and (iv) LN.*

Proof appears in Appendix B. Depending on exactly where t_4 lies, there are three different combinations of the thresholds that support the observed data.[25]

[25]These are: $t_1^M < t_3^M < t_2^M < t_4^M$, $t_1^M < t_3^M < t_4^M < t_2^M$, and $t_1^M < t_4^M < t_3^M < t_2^M$. We are not able to distinguish among the three cases without making additional assumptions or without more detailed data, as they all imply the same ranking.

Similarly, Observation 2 establishes the ranking among party donors which differs from members in one fundamental aspect: low valence donors are placed, on average, in better ranked positions than high valence non-donors; the opposite is true for party members. The following proposition argues that there is only one order of the thresholds that can generate the observed ranking among donors. Proof appears in Appendix B.

Proposition 2 (Donations)**.** *Consider political donations as a measure of intra-party value. If and only if* $t_4^D < t_1^D < t_2^D < t_3^D$*, the group ranking is as follows: (i) HD; (ii) LD; (iii) HND; and (iv) LND.*

The model used here enables us to understand the ranking differences between members and donors. First, note that among donors $t_2^D < t_3^D$, while the opposite is true among members $t_3^M < t_2^M$. Since t_3 is the threshold below which high valence candidates are willing to run, it is the same in both cases, so $t_3^M = t_3^D = t_3$, which implies that $t_2^D < t_3 < t_2^M$. Therefore, donors must be rewarded with better slate positions than members in order to meet their participation constraints. In other words, donation is more costly than membership ($c^D > c^M$). Second, the value of donors to the party leader exceeds the value of being a party member. This follows from the fact that for members $t_1^M < t_4^M$, whereas for donors $t_4^D < t_1^D$. These two facts along with the cost differences described earlier, $c^D > c^M$, yield that $t_4^D < t_1^D < t_1^M < t_4^M$ implying that $\gamma^D > \gamma^M$. Proposition 3 summarizes both implications.

Proposition 3 (Comparison)**.** *Suppose the slate is ranked as proposed in Observations 1 and 2. Then, our theoretical framework predicts that for candidates, becoming a donor is more costly than becoming a member,* $c^D > c^M$*, and for party leaders, donors are more valuable than members of the same valence,* $\gamma^D > \gamma^M$*.*

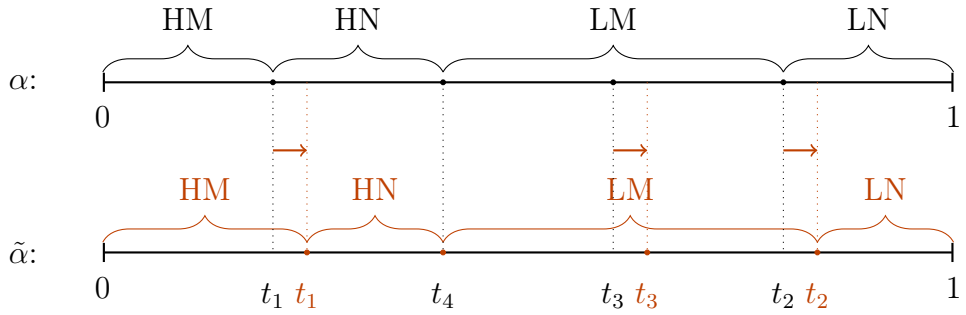
The model can therefore rationalize the reversal in ranking between party donors and members by donations being more costly for candidates and more valuable for political parties, which we view as reasonable and intuitive. Finally, the model is also consistent with Observation 5. In particular, the model predicts that an increase in popularity leads to a higher share of high valence candidates with high intra-party value and a decrease in low valence candidates with low intra-party value, which is what we find in the data for both members and donors.

Proposition 4 (Strength of Parties)**.** *An increase in the popularity of a party represented by an increase in* α *leads to a higher share of high valence candidates with high intra-party value and a lower share of low valence candidates with low intra-party value on the slate.*

Proposition 4 follows from relaxing the participation constraints of all candidates. As α increases, so does the probability of being elected at any slate rank, *ceteris paribus*. The changes in the shares of the two remaining types of candidates are generally ambiguous and depend on the relative shifts of different thresholds. The thresholds are complex to characterize, as they depend on several features including the slope of the probability function, and the relative shifts are therefore not easy to calculate. The suggested ranking for donors stated in Proposition 2 additionally implies that the overall share of donors, both with high and low valence, always rises when α increases. For members, since there are several possible combinations, not much more can be said about the two middle groups of candidates.

To provide intuition, consider one particular combination of thresholds: $t_1 < t_4 < t_3 < t_2$.[26] As a party experiences a positive popularity shock, an increase in α to $\tilde{\alpha} > \alpha$, the participation constraints relax for all types of candidates. This shifts t_1, t_2, and t_3 towards the bottom of the slate as displayed in Figure 1.4. Since t_4 does not change, the shares of high and low valence candidates remain unchanged, but the share of members (high and low valence together) increases.

Figure 1.4: Explaining Membership Data



1.5 Concluding Remarks

We approach the process of the selection of political candidates in PR systems as a market. On the one hand, a party leader (the demand side) demands valence and intra-party value in exchange for slate positions that are more likely to win a mandate. On the other hand, candidates (the supply side) decide on their intra-party value, as they strive to

[26] This is our preferred combination as it unambiguously predicts an increase in members in response to a positive party shock, which is the most pronounced effect that we found in the data.

win a mandate on a municipal council. This interaction resembles typical market forces. We support the market-like interpretation by empirical evidence. First, candidates are ranked as predicted by market mechanisms in which the top positions tend to be occupied by high valence candidates with intra-party value, whereas the bottom positions tend to be occupied by the candidates that are the least valuable. Second, we document that party leaders seem to face a trade-off between candidates' intra-party value and their attractiveness to voters. Third, with increasing popularity of the party, the party leader takes advantage of her position to form a slate with a higher intra-party value, as she has more to offer to candidates in exchange for their intra-party value. Fourth, higher intra-party value (more costly intra-party action) tends to be rewarded by better slate positions. This follows from: (i) a comparison between party membership and, arguably more costly, party donations; and (ii) the positive link between the size of donations and better slate position. Systematic ranking of candidates has an important methodological implication. The fact that high valence candidates and candidates with high intra-party value are over-represented in positions with higher probability of being elected casts doubt on the frequently used approach that evaluates a slate by considering the simple shares of different groups of candidates on the slate rather than considering their distribution on the slate. In fact, this approach may easily lead to misleading results, even in (semi-) open list electoral systems.

The gate-keeping power of parties is likely to give rise to a principal-agent problem in which party leaders may pursue their private goals in political selection. Swing voters incentivize the party leader to care about valence, which mitigates the problem, assuming that voters' concern is candidates' valence. The interests of the party leader and voters are aligned at the top positions where high valence candidates are willing to increase their intra-party value. The conflict between a party leader's interests and the interests of the public tends to appear at the lower slate positions where the party leader has the opportunity to skew the selection and ranking of the candidates in her favor, by prioritizing low valence candidates with intra-party value rather than high valence candidates with no intra-party value.

We assign candidates two characteristics (valence and intra-party value) and relax an assumption which is prevalent in the literature. While this mitigates the principal agent problem, it may intensify other problems such as rent seeking. If being of high valence does not guarantee that candidates will be placed in well ranked slate positions, everyone is incentivized to acquire more intra-party value, which may take different forms and may

not be limited to membership status and political donations. In fact, intra-party value can be a very broad concept that can include a wide variety of attributes. For example, employees of the party, public proponents or anyone providing services of any kind to the party may be considered of high intra-party value, regardless of whether they are also members or donors. More importantly, any rent seeking activity that a candidate engages in for the benefit of the political party may be seen by the party leader as increasing his value to the party.

While this chapter describes the process of selecting and ordering candidates on a slate as a trade between party leaders and candidates, it is mute about the exact mechanisms. It does not address the structure of the market, nor the forms of contracts between candidates and parties. As candidates and party leaders interact in highly uncertain environments and contracts between them are potentially dynamic, there are other possible research questions to study. For example, who bears the cost of uncertainty? Do candidates in marginal positions make donations prior to an election or only after being elected? Do party leaders enforce party affiliation after the election and does such enforcement depend on the valence of candidates? Furthermore, this chapter has not addressed interactions among different political parties within a municipality, but future research may shed light on the influence of political competition on the interaction of parties and candidates.

1.6 Appendix A

Appendix A is arranged so it matches the structure of the empirical sections of the chapter. First, it provides additional descriptive statistics regarding the dataset used. Second, it defines *Rank*. Third, it provides additional robustness exercises confirming that the patterns described in Section 1.3 remain unchanged in several alternative settings. Fourth, we analyze the effect of party popularity on the slate structure as conducted in Section 1.3.2 in more detail by estimating it separately for each political party

Table 1.6: Share of Candidates by Types of Ballots

Type of Ballot	2002	2006	2010	2014	2018
Main Parties	43.5%	41.6%	39.0%	35.4%	29.3%
Local Parties	20.7%	18.2%	20.6%	23.5%	26.7%
Independent candidates (> 1)	29.2%	34.6%	35.7%	36.9%	39.6%
Independent candidate (solo)	6.6%	5.6%	4.8%	4.4%	4.4%

The table summarizes shares of candidates running on different types of ballots.

Data Description

Table 1.7 shows a distribution of candidates according to the parties under which they run. As we describe in Section 1.2, only candidates running on slates of one of the six main political parties are kept in the data; our main results are based on 5 years of municipal elections: 2002, 2006, 2010, 2014, and 2018. Data from 1998 are used only to control for the previous political experience of candidates. Note that two parties, TOP 09 and ANO, participated in only three and two elections, respectively. The drop in the number of candidates running on the 2018 TOP 09 slate is because we exclude candidates and slates when there was a joint slate of more parties, which was the approach of TOP 09 in most municipalities in 2018. Table 1.8 shows shares of formal members on parties' slates. There is a significant variation both in time and across parties. The recently established parties, TOP 09 and ANO, tend to have fewer members on the slates.

Measure of performance of candidates To compare candidates' performance we measure candidates' relative performance within his slate $Relative\,Votes$ as a ratio of his votes and the average share of of votes cast to a candidate listed on the slate. Receiving the average number of votes of one's party means that the $Relative\,Votes$ equals 1.

Table 1.7: Number of Candidates

Political Party	2002	2006	2010	2014	2018	Total
KDUCSL	17,717	17,930	14,940	14,603	12,238	77,428
CSSD	16,095	16,111	16,884	16,336	11,752	77,178
KSCM	20,717	19,074	17,375	16,083	12,704	85,953
ODS	16,168	19,042	18,757	11,667	10,615	76,249
TOP 09	0	0	9,703	6,363	1,338	17,404
ANO	0	0	0	7,906	7,927	15,833
Total	70,697	72,157	77,659	72,958	56,574	350,045

Table 1.8: Share of Party Members

Political Party	2002	2006	2010	2014	2018	Average
KDUCSL	37%	34%	31%	27%	27%	31.2%
CSSD	43%	41%	48%	50%	50%	46.4%
KSCM	60%	55%	52%	48%	48%	52.6%
ODS	48%	51%	51%	50%	43%	48.6%
TOP 09	.	.	27%	29%	35%	30.3%
ANO	.	.	.	18%	27%	22.5%
Average	47.0%	45.3%	42.8%	37.0%	38.3%	

Figure 1.5 shows the average *Relative Votes* among candidates listed on slates with 15 candidates.

Figure 1.5: Average Relative Votes for Different Rank on Slate with 15 Candidates

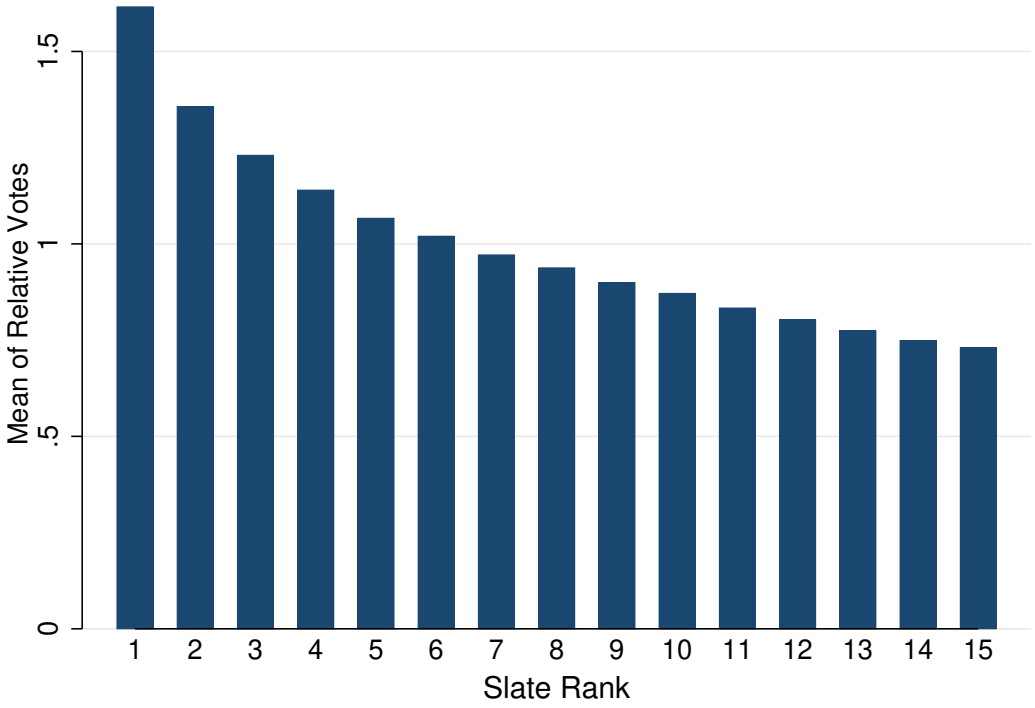
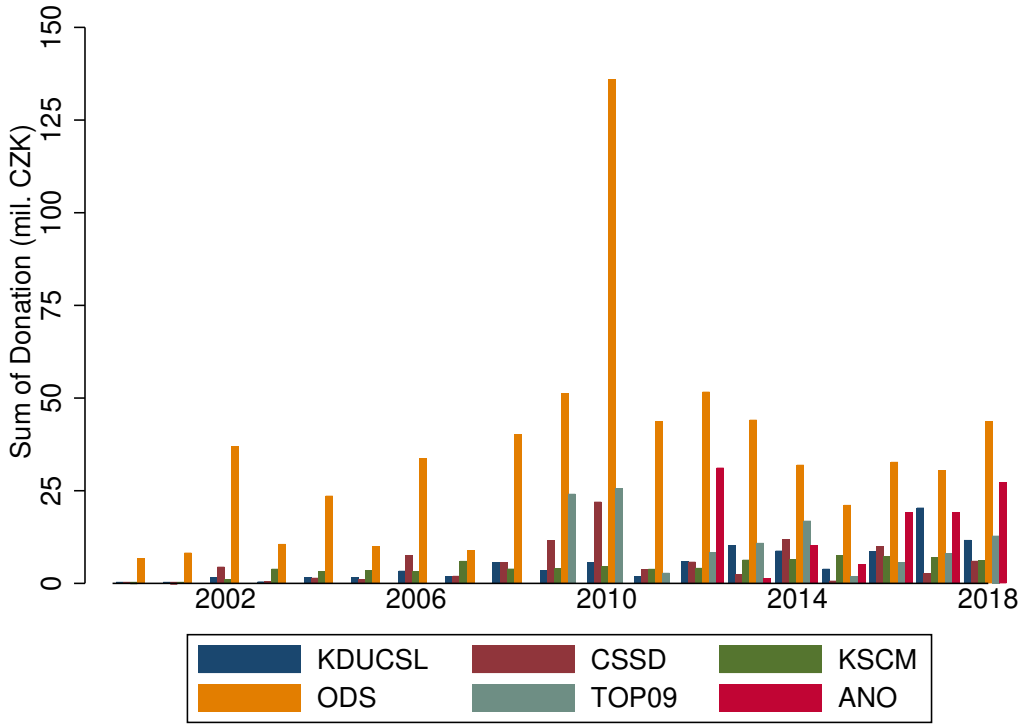


Figure 1.6: Total Donation Made by Candidates



Notes: Only donations made by the party's candidates in any municipal elections are included.

Definition of (Normalized) Rank

The number of candidates on slates differs across municipalities, parties, and election years. To provide a reasonable comparison with respect to the main characteristics of our interest, our preferable variable *Rank* is conditional on other observable characteristics of candidates and normalized to be between 0 and 1. In the first step, we define *RelativeOrderCandidate* as follows

$$RelativeOrderCandidate = \frac{\text{position on slate} - 1}{\text{number of candidates on a slate} - 1}. \tag{1.9}$$

Having defined *RelativeOrderCandidate*, we run the following regression

$$\begin{aligned} RelativeOrderCandidate = & \sum_{i \in \{AgeGroup\}} \beta_i^1 dm\_age_i + \sum_{i \in \{MunElections\}} \beta_i^2 dm\_MunElection_i + \\ & + \beta^3 dm\_Male + \sum_{i \in \{PolParty\}} \beta_i^4 \ dm\_PolParty_i + \sum_{i \in \{PrevMunMan\}} \beta_i^5 \ dm\_PreMuniMan \\ & + \sum_{i \in \{PrevMunCandidates\}} \beta_i^6 \ dm\_PrevMunCandidates + \varepsilon \end{aligned} \tag{1.10}$$

In the second step, we take the residuals $\hat{\varepsilon}$ from regression 1.10 and normalized them to fall in interval (0,1).

$$Rank_i = \frac{\hat{\varepsilon}_i - \hat{\varepsilon}_{min}}{\hat{\varepsilon}_{max} - \hat{\varepsilon}_{min}} \tag{1.11}$$

To provide more robust evidence, we replicate our results using an unconditional measure of rank. Similarly to the main measure, the alternative is normalized to be within an interval 0 and 1. Conversely, it is not derived from residuals of a regression, but directly from candidates' positions. The top position has *Rank* of 0, while the bottom position of 1. The interior slate positions map to different *Rank* values depending on the number of candidates on the slate. Suppose a candidate i is placed on the k-th position on a slate with n candidates. Then the *Normalized Rank* is defined as follows:

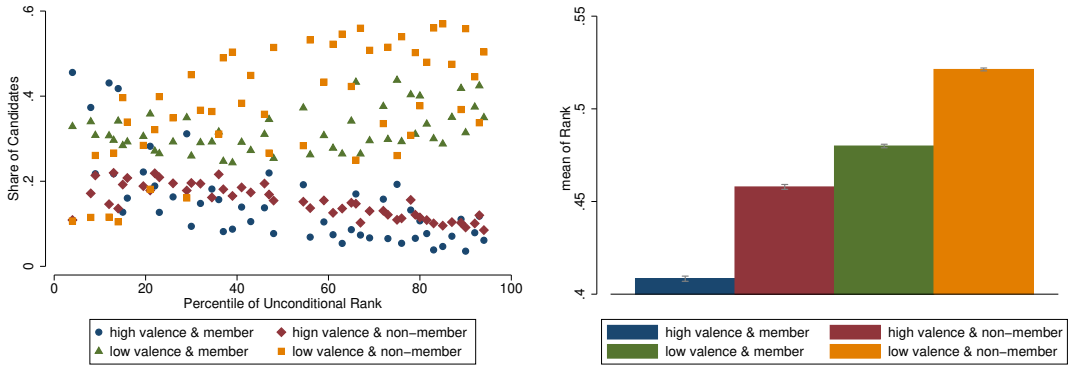
$$Normalized\ Rank = \frac{k-1}{n-1}. \tag{1.12}$$

Slate Structure - Robustness Exercise

We provide two types of robustness exercises. First, we replicate graphs from Section 1.3.2 using *Normalized Rank* instead of *Rank* as in the main analysis and show that the ranking patterns remain unchanged. Second, we demonstrate robustness of results from Section 1.3.2 by running the same exercises for differently specified samples.

Normalized Rank In Section 1.3.2, we use *Rank* in our analysis in order to separate away the effects of age, gender, political experience and other characteristics and we focus only on the effect of valence and intra-party value. Though we consider this methodology more informative, the interpretation of the results is complicated. We therefore replicate the figures from Section 1.3.2 using *NormalizedRank*. The figures show that the ranking patterns remain unchanged.

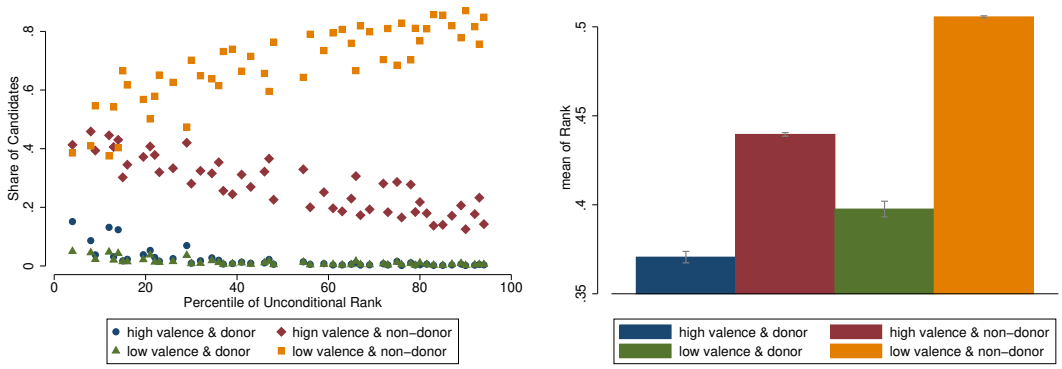
Figure 1.7: Slate Structure for Members with *Normalized Rank*



(a) Slate Structure (Members) **(b)** Average Rank by Types (Members)

Slates With All Types of Candidates We next provide additional exercises to demonstrate that Observations 1 and 2 persist under different circumstances. First, we study only candidates who run on slates that include all types of candidates. In other words, if one or more of the groups (HM, HN, LM, LN) is missing from the slate, every candidate on the slate is excluded from our analysis in this exercise. Figures 1.9a and 1.9b show that the ranking patterns remain very similar and are even more pronounced than the patterns in Section 1.3 which is consistent with our suggested mechanisms. Compared to the baseline figures, the number of candidates drops to roughly half. Interestingly, the share of low valence non-members placed in well-ranked positions drops significantly, as a

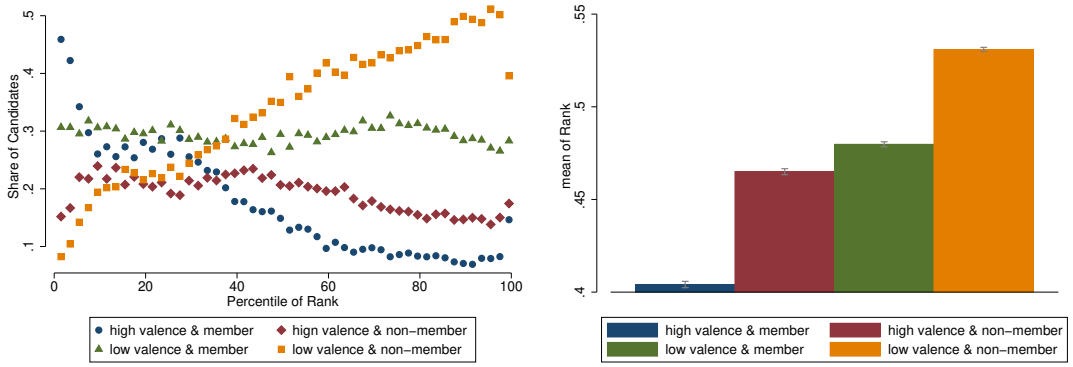
Figure 1.8: Slate Structure for Donors with *Normalized Rank*



(a) Slate Structure (Donors) **(b)** Average Rank by Types (Donors)

sizable portion of low valence non-members are well-placed on slates that do not include all four types of candidates.

Figure 1.9: Slates with All Four Groups of Candidates

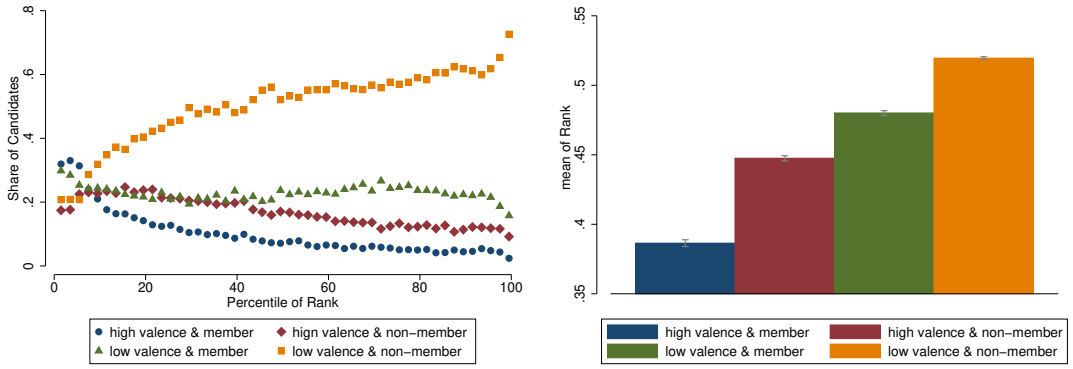


(a) Slate Structure (Members) **(b)** Average Rank by Types (Members)

Novice Candidates Second, we study only a subset of candidates who run in a municipality for the first time and thus have no prior experience in municipal elections. This specification is likely to be robust against different forms of historical relationships between candidates and the party. Figures 1.10a and 1.10b show that the ranking patterns hold for political novices as well. Interestingly, there is no peak at the bottom of the slate, suggesting that the peak is indeed driven by politically experienced candidates.

Party Donors in Parliamentary Elections Third, we provide additional evidence from parliamentary elections. The share of party donors among candidates in municipal

Figure 1.10: Only Candidates Without Prior Political Experience



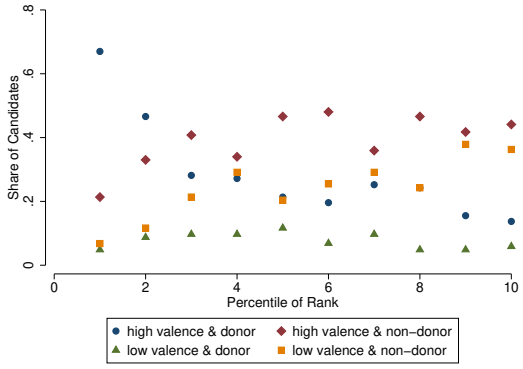
(a) Slate Structure (Members) **(b)** Average Rank by Types (Members)

elections is relatively small, as is the variation across candidates in different subgroups. Therefore, to provide additional evidence of sorting on the slate among party donors, we study slates in parliamentary elections. While the number of candidates from one of the six main parties in the last 5 parliamentary elections is *only* around 8,500, roughly a third are classified as party donors. For the first exercise, a candidate is classified as a donor if his or her donation was any positive number, including small amounts. We create *Rank* as before, normalizing the slate position to be within the [0,1] interval.

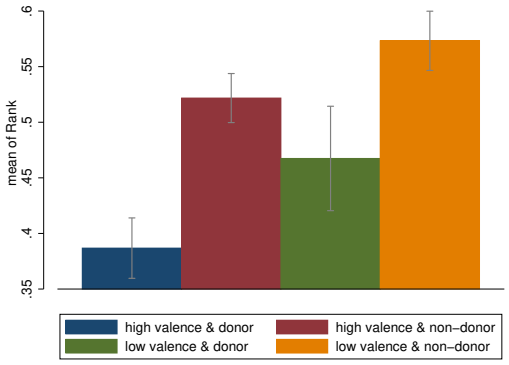
Figure 1.11a collapses candidates according to their rank into ten 10% bins. The share of high valence donors decreases rapidly as one moves to worse ranked positions on the slate. While there are almost two thirds of high valence donors among the 10% best ranked positions, there are only around 15% among the worst ranked candidates. Importantly, similarly to municipal elections, in parliamentary elections, low valence donors are ranked better than high valence non-donors. We thus obtain qualitatively the same observation for both municipal election slates and parliamentary election slates.

Next, we reclassify the group of donors to those who donate at least CZK 50,000 (approx. EUR 2,000). Figures 1.12a and 1.12b show the slate structure for more generous donors. In line with the model presented, as the intra-party value becomes more costly to the candidates and more valuable to the party, the share of donors shrinks, while their slate rankings improve. In fact, as the threshold for donors rises, the differences in ranking between high and low valence donors disappears.

Figure 1.11: Slates in Parliamentary Election (Donors)

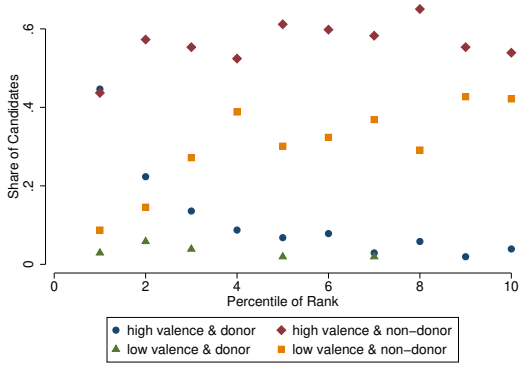


(a) Slate Structure (Donors)

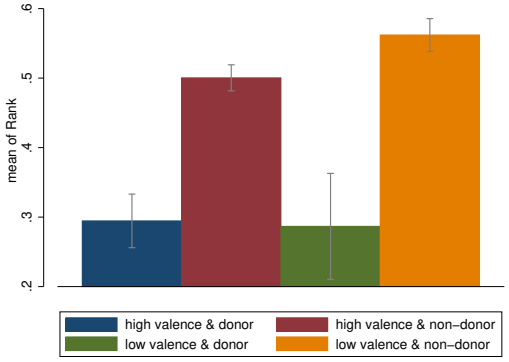


(b) Average Rank by Groups (Donors)

Figure 1.12: Slates in Parliamentary Election (Generous Donors)



(a) Slate Structure (Generous Donors)



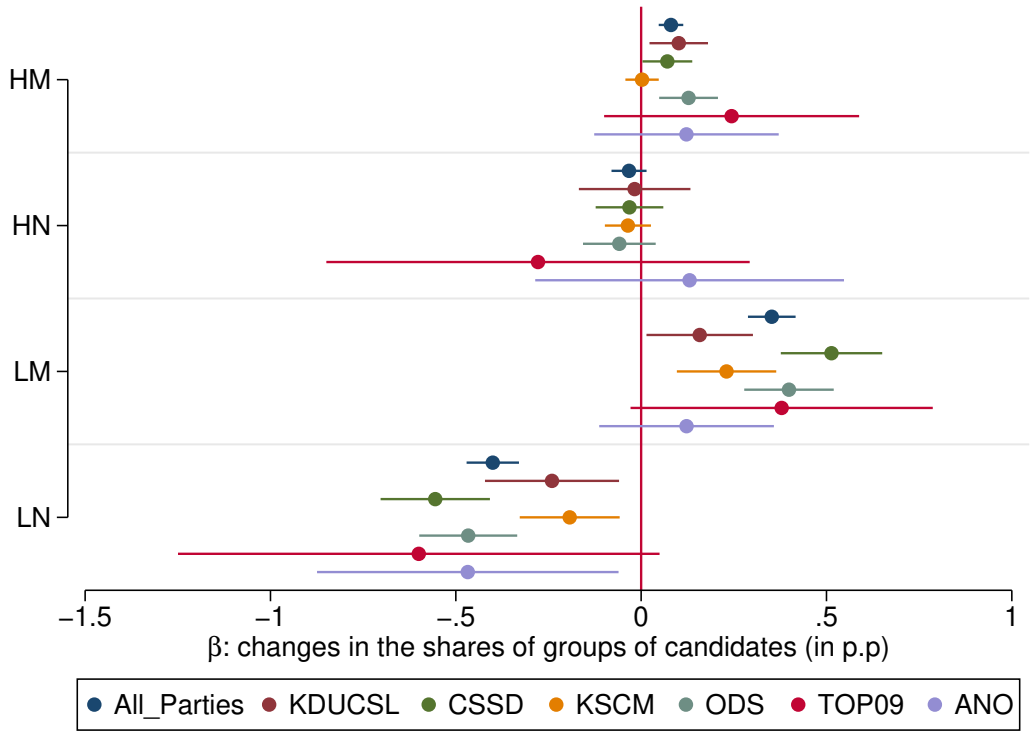
(b) Average Rank by Groups (Generous Donors)

Effect of Popularity of Party on Slate Structure

We also extend the results showing that after a popularity shock, there is an increase in shares of high valence candidates and candidates with high intra-party value. In particular, we show that the effect is comparable across all political parties studied and thus is not driven by one party, and we show the same results using a different source of variation.

Party Heterogeneity Figure 1.13 shows changes in slate structure after a popularity shock, decomposed for all six parties. It shows that a positive popularity shock is followed by a weak increase in high valence members in all parties. The share of low valence members increases as well. The predicted drop in the share of low valence non-members is also prevalent among all parties. Note that for both TOP09 and ANO, the coefficients have relatively large confidence intervals, as the parties have participated only in three and two elections, respectively, and thus the estimates are based on fewer observations.

Figure 1.13: Changes in Group Shares (Members) by Party



Different Sources of Variation To provide additional evidence supporting our narrative, we explore different sources of variations in party power. Specifically, compared to

the baseline specification as in Regression 1.3, we employ two different fixed effects: fixed effects for party-municipality cell (γ_{pj}) as before; and fixed effect for election (γ_τ). See regression 1.13. Therefore, we do not control for variation caused by a change in party popularity at the national level. Suppose party A becomes more popular; this popularity shock increases both the share of votes in a national election in the municipality and the party's electoral potential in the next municipal election.

$$Share^g_{pj\tau} = \alpha^g + \beta^g PE\ ShareVotes_{pj\tau} + \sum_{k\in\{HM,HN,LM\}} \delta^k PE\ Share^k_{p\tilde{j}\tau} + \gamma^g_{pj} + \gamma^g_\tau + \epsilon^g_{pj\tau} \tag{1.13}$$

Figure 1.14: Changes in Group Shares on the Slate

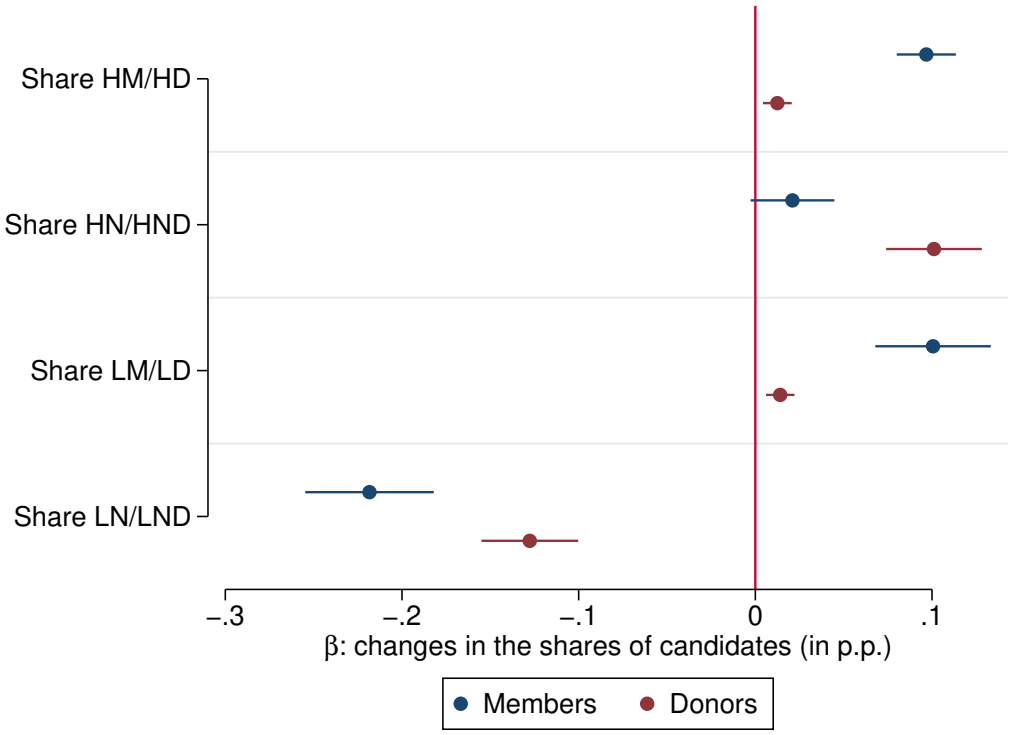


Figure 1.14 graphically shows coefficients β^g for both measures of intra-party values. The main narratives hold. As a party becomes more popular and expects more mandates to win, there are more high valence candidates and more candidates with high intra-party value on the slate. Consequently, the least valuable group, low valence candidates with low intra-party value, are forced out.

1.7 Appendix B

Optimal t_4

Lemma 1. *Suppose t_1, t_2, and $t_3 \in [0,1]$. Then t_4 implicitly defined as $g(t_4) = \gamma^a$ is a solution to the party leader's problem. Formally,*

$$t_4 \in \underset{\tilde{t}}{\text{argmax}} \; V(\bar{v}(\tilde{t}), \bar{m}(\tilde{t}) | t_1, t_2, t_3) \tag{1.14}$$

If $t_4 < min(t_2, t_3)$ then t_4 is a unique solution of the party leader's problem.

$$t_4 = \underset{\tilde{t}}{\text{argmax}} \; V(\bar{v}(\tilde{t}), \bar{m}(\tilde{t}) | t_1, t_2, t_3) \tag{1.15}$$

Proof. To see this, we will solve the party leader's problem. To fix the notation, we use the membership notation for the measure of intra-party value. The party leader chooses a threshold $\tilde{t}$, such that it maximizes her objective function $V(\bar{v}, \bar{m})$:

$$\max_{\tilde{t}} V(\bar{v}, \bar{m}) = \max_{\tilde{t}} \int_{HM} g(t)dt + \int_{HM} \gamma^a dt + \int_{HN} g(t)dt + \int_{LM} \gamma^a dt \tag{1.16}$$

The first two terms of the objective function represent the valence andintra-party value of high valence members and are independent of the party leader's choice of $\tilde{t}$. That simplifies the problem into a sum of two integrals.

$$\max_{\tilde{t}} \tilde{V} = \max_{\tilde{t}} \int_{HN} g(t)dt + \int_{LM} \gamma^a dt \tag{1.17}$$

Remember that t_2 and t_3 are the worst positions from which LM and HN are willing to run, respectively. The only trade-off for the party leader occurs for positions in which both these groups of candidates are willing to run. Therefore, for $\tilde{t} > \min\{t_2, t_3\}$ there is no trade-off and any choice of $\tilde{t}$ maximizes the objective function.

If $\tilde{t} < \min\{t_2, t_3\}$ then the problem appears as follows

$$\max_{\tilde{t}} \tilde{V} = \max_{\tilde{t}} \int_{t_1}^{\tilde{t}} g(t)dt + \int_{\tilde{t}}^{\min(t_2,t_3)} \gamma^a dt, \tag{1.18}$$

Deriving the first order conditions and denoting the solution as t_4 yields

$$g(t_4) = \gamma^a. \tag{1.19}$$

□

Proofs of Propositions

We prove Proposition 1 and 2 simultaneously by considering all possible combinations of thresholds and the associated orders of groups of candidates.

As there are four different thresholds t_1, t_2, t_3, and t_4 ordered on a continuous interval $[0, 1]$, there are 24 different combinations in which they may be ordered. First, note that it must be the case that $t_1 < t_3$, otherwise intra-party value would impose a negative cost, i.e. $c^a < 0$. Similarly, it must be the case that $t_1 < t_2$, otherwise running would impose a negative cost for high valence candidates, i.e. $c^h < 0$. That leaves eight possible cases.

Second, note that if all four groups are represented on a slate, it must be the case that $t_2 > \min\{t_3, t_4\}$. Suppose the opposite is true and $t_2 < t_4$ & $t_2 < t_3$, then low valence candidates with high intra-party value (LM candidates) will be willing to run only in positions for which high valence candidates with low intra-party value are preferable and willing to run. Therefore, LM would not be represented on the slate. That excludes an additional two combinations.

We are left with six combinations of thresholds. Note that four thresholds divide the slate into five intervals. We next describe which types of candidates (using a notation for membership status) will be in which intervals.

$t_1 < t_3 < t_2 < t_4$ implies the following intervals {HM, HN, LM, LN, LN}

$t_1 < t_3 < t_4 < t_2$ implies the following intervals {HM, HN, LM, LN, LN}

$t_1 < t_4 < t_2 < t_3$ implies the following intervals {HM, HN, LM, HN, LM}

$t_1 < t_4 < t_3 < t_2$ implies the following intervals {HM, HN, LM, LM, LN}

$t_4 < t_1 < t_2 < t_3$ implies the following intervals {HM, HM, LM, HN, LN}

$t_4 < t_1 < t_3 < t_2$ implies the following intervals {HM, HM, LM, LM, LN}

Note that HN are missing in (f). Case (c) is a special case, as HN occupy two disconnected intervals. If this were true, we should observe high variance in HN candidates' positions, which is not the case. Therefore, we rule the case (c) out as not representing the data.

Finally, the case (e) is the only possible case that implies that the average position of low valence candidates with high intra-party value is *better* than the average position of high valence candidates with low intra-party value. That proves Proposition 2. Cases (a), (b), and (d) are the only three cases that: (i) satisfy the conditions from Proposition 1 ($t_1 < t_3 < t_2$ & $t_1 < t_4$); and at the same time: (ii) imply the sorting of candidates observed in the data. This proves Proposition 1.

Proposition 4 follows by looking at the threshold orderings and shifting t_1, t_2 and t_3 to the right. However much they shift, the HM interval always increases and the LN interval is always reduced. We omitted cases where $t_1 < 0$ or does not exist and HM are not present. In such case, the share of the group at the top of the slate increases instead.

Chapter 2

Sentencing Decisions Around Quantity Thresholds: Theory and Experiment

This chapter was coauthored with Jakub Drápal (Faculty of Law, Charles University).

2.1 Introduction

Scholars have been studying sentencing shortcomings particularly since Frankel (1972) described sentencing as lawlessness. Sentencing disparities - treating alike cases differently and different cases alike - have become one of their primary focuses. Since then, scholars have identified various characteristics of judges, offenders, and victims that contribute to sentencing disparities (Sporer and Goodman-Delahunty, 2009a). Ulmer (2012) and Ulmer and Bradley (2019) provide a comprehensive summary of the literature on sentencing disparities.

To render sentencing more consistent and principled, criminal justice systems around the world have introduced various measures (Council of Europe, 1992; Clarkson and Morgan, 1995; Ashworth, 2009). Some of the measures, however, create new sources of disparities. For example, while the U.S. federal sentencing guidelines reduced the level of sentencing disparities at the court level (Anderson et al., 1999), they introduced unwarranted disparities at other levels. Hofer (2019) documents that the guidelines with mandatory minimums lead to higher racial disparities. Furthermore, as prosecutors have gained more discretion, disparities have been further displaced to earlier stages of the criminal process, resulting in large and unjustifiable trial tax (Johnson, 2019) and in

charge disparities (Shermer and Johnson, 2010; Tuttle, 2019). While judge consistency may have improved, the overall effect of guidelines is considered less successful. Some even argue that the U.S. federal sentencing guidelines need to be repudiated (Tonry, 2019). The U.S. federal sentencing guidelines serve as a perfect example of how efforts to reduce sentencing disparities may unintentionally cause disparities.

In this chapter, we identify a new source of sentencing disparities resulting from one of the oldest measures aimed to structure sentencing discretion: Offense subsections with specific sentencing ranges. Since the same offense can vary significantly in its circumstances and seriousness, many legal systems divide offenses into subsections of more or less serious behavior with specific, usually overlapping, sentencing ranges. Such provisions are a common measure structuring sentencers'[1] discretion, especially in Europe. A finer structure of offense subsections raise the question of how to determine thresholds dividing offenses into subsections. To classify cases into a corresponding subsection, criminal codes [2] often rely on measurable and quantifiable variables, such as caused damage and amount of drug possessed, using so called quantity thresholds (Foulds and Nutt, 2020; Bjerk, 2017a). Figure 2.1 represents an example of a structure of offenses studied in this chapter.

We argue that quantity thresholds are likely to introduce a new form of sentencing disparities. To study the consequences of thresholds on sentencing decisions formally, we develop a simplified theory assuming that sentencers impose a sentence within a sentencing range according to the seriousness of the particular case relative to other cases sentenced within the same subsection. The proposed theory implies that thresholds can cause a sizeable sentencing disparity. Furthermore, thresholds can also lead to perverse sentencing, when more serious cases are systematically sentenced to a more lenient punishment.

The overall effect of a threshold on sentencing decisions can be decomposed into two opposing effects: the *severity* and the *reference* effects. The proposed theory allows us to describe these effects and discuss their properties. Consider two thefts that marginally differ in the amount of damage. One case - say A - is just below a threshold, whereas the other one - say B - is just above a threshold. The sentencer's considerations in these two cases change in two aspects. On the one hand, case B is sentenced according to a

[1] For the sake of simplicity, when applicable, we refer to both judges and prosecutors as sentencers.

[2] Throughout the text we talk about the criminal code, yet most of our conclusions apply to sentencing guidelines as well.

Figure 2.1: Theft Subsections with Sentencing Ranges

Notes: This figure represents the studied problem using an example of theft. The theft offense in the Czech Republic was divided into four subsections by strict thresholds based on caused damage (CZK 50,000, CZK 500,000, and CZK 5,000,000; from October 2020, several months after our experiment, the thresholds doubled). For each subsection, there is a specific sentencing range that determines the lower and the upper length of incarceration. In many cases, the sentencing ranges overlap and thus effectively permit perverse sentencing when a (marginally) more serious case is sentenced to a more lenient punishment.

higher sentencing range. We call the change in a sentence evoked by this mechanism the *severity* effect and argue that it leads to harsher sentences. On the other hand, case B is compared to arguably more serious cases. As a result, the relative position within the sentencing range is likely to be lower. We call it the *reference* effect and argue that it tends to decrease the sentence. Depending on which effect dominates, the effect of the threshold is either positive – case B is sentenced more harshly –, or negative – case B is sentenced less harshly.

To provide empirical evidence, we conduct a Rachlinski-style online experiment with 200 Czech prosecutors. Each participant of the experiment was presented with two vignettes, each describing a different crime case: theft and drug possession. In both cases, we implement several scenarios that vary an amount of classifying variables (amount of methamphetamine possessed and caused damage) around thresholds. We then use the

variation to estimate the causal effects of thresholds. Additionally, to test for the existence of the severity and the reference mechanisms, we implement two additional scenarios that introduce an isolated variation in the composition of cases in an offense subsection and sentencing ranges, respectively.

The causal effect of thresholds on sentencing decisions is enormous and leads to a sizeable increase in sentences. In the theft vignette, we vary the amount of damage in four scenarios around two thresholds of CZK 50,000 and CZK 500,000 (henceforth the 50k and the 500k thresholds, respectively). To estimate the effect of the 50k threshold, we compare identical cases with marginally different damage of CZK 48,283 and of CZK 51,283. The 50k threshold increases the average sentence by more than 10 months which represents a 50% increase. Looking at the effect of the 500k threshold, we compare identical cases with damage of CZK 487,092 and CZK 508,213 and estimate the effect to be around 4 months, which corresponds to 10%. Interestingly, the increase in the average sentence caused by the 50k threshold is not statistically different from an increase caused by increasing damage within the same subsection from CZK 51,283 to CZK 487,092, i.e. by approx. CZK 435,000. This demonstrates the enormous size of the 50k threshold effect.

In the drug possession case, the effect of the threshold is also of a great magnitude. The average recommended sentence in a case with 147.8 g of methamphetamine was 24.4 months, while in the identical case with 151.8 g of methamphetamine it was 31 months. The 6month increase (25%) is in stark contrast to the 3% increase in the amount of drug possessed. We next make two observations about the severity and the reference mechanisms. First, the fact that the overall threshold effect is positive suggests that the severity mechanism dominates the reference one. Second, the test for the reference mechanism provides suggestive evidence that it is negative as predicted.

Finally, to provide more comprehensive evidence on sentencing disparities caused by thresholds, we propose a novel parametric measure of (in)justice. Consider two cases and corresponding sentences A and B and suppose that case B is marginally more serious. We define sentence B to be *just*, if it is not more lenient than sentence A and not too (unreasonably) harsh compared to sentence A. Consequently, there are two reasons why sentence B might not be considered just: (i) sentence B is more lenient than sentence A (*type I injustice*); and (ii) sentence B is too harsh (*type II injustice*). Using these definitions, we quantify the shares of just decisions in our experiment. The results generally follow the ones of the average recommended sentence suggesting that, due to thresholds,

prosecutors view and treat almost identical cases differently.

Our simplified theory does not allow us to predict the exact magnitude of the effects or to provide more detailed discussion regarding the heterogeneinty of the effect. To do so, we would need more strict assumptions of how exactly sentencers arrange cases within the sentencing range. However, we still can derive interesting implications. The reference effect is a linear combination of an increase in the lower limit and in the upper limit, with the relative position of the case in the sentencing range as a weighting coefficient. Therefore, relatively more serious cases are more sensitive to an increase in the upper limit, while less serious cases to an increase in the lower limit.

The rest of the chapter is organized as follows. We first propose the theory that guides the experimental design. We then introduce the experiment and discuss the results. In particular, we focus on the treatment effect on the average recommended sentence. Next, we introduce the novel measures of (in)justice and apply these measures on the experimental data. Finally, we discuss the limitations of our project and implications for future research.

2.2 Theory

2.2.1 Offenses and Thresholds that Divide Them into Subsections

Criminal codes categorize offenses and corresponding sentencing ranges in different ways. In most systems, offenses are divided into subsections with specific sentencing ranges according to certain factors. A finer structure of offenses narrows down sentencers' discretion and provides them with guidance. To divide offenses, criminal codes rely on various factors, some of which are quantifiable such as the amount of damage[3], amount of drugs[4], and number of days when the victim of an assault was not able to work.[5] Quantity thresholds have existed for centuries. For example, theft was divided into subsections with specific sentencing ranges by stolen amount at least since 1803, when the Criminal

[3]Found e.g. in English and Welsh sentencing guidelines for theft or fraud or in Russian (The Criminal Code of the Russian Federation, no. 63-FZ of 13. 6. 1996, note in Article 146) or Czech criminal code (law n. 40/2009 Coll., Section 138).

[4]Set e.g. by the United States Federal Sentencing Guidelines (Drug Quantity Table in Section §2D1.1.) and in Norway (Norway establishes a specific sentencing range for offenses involving a very substantial amount, see e.g. Sections 232 and 233 of the Norwegian Criminal Code) or Slovak Criminal Code (law n. 300/2005 Coll, s. 135

[5]E.g. in Slovakia (Slovak Criminal Code s. 123/4) and the Czech Republic (Decision of the Czechoslovak Supreme Court n. Tpjf 24/85 [R 16/1986 tr.])

Code of the Austrian Empire set thresholds for thefts at 25 and 300 ducats.[6] The typical domain of quantity thresholds, however, are drug related offenses (Leader-Elliott, 2012; Sentencing Council, 2011). Despite their popularity, many view them as problematic. In particular, Fleetwood (2011) argues that factors such as the role in drug-trafficking capture offense seriousness better than drug amounts. Furthermore, the amount of drug may be easily manipulated, even by law enforcement officers (Travova, 2019).

Implementation of offense subsections and corresponding sentencing ranges varies across criminal justice systems. Some criminal justice systems vary only the upper limits of sentencing ranges and do not specify the lower limits (e.g. France), whereas others set both the lower and the upper limits for each subsection individually. Additionally, criminal justice systems differ in sentencers' discretion to impose a sentence below and above the prescribed sentencing range (Kert et al., 2015; Kaspar, 2020).

While specific sentencing ranges for offense subsections divided by quantity thresholds provide guidance to sentencers, they may also introduce a new source of disparities. Virtually identical cases – such as, e.g., a theft of 299 (case A) and 300 (case B) ducats – fall in two subsections with different sentencing ranges, resulting in possibly different sentences for A and B. Sentencing ranges serve as a rudimentary signpost indicating to what extent severe sanctions are expected. The composition of cases within a corresponding subsection, defined by a quantity threshold, provide a natural reference group within which criminal acts are compared with each other and are ordered by their seriousness. Using the terminology of von Hirsch (2017), the sentencing ranges serve as cardinal and the subsection composition as ordinal guidance for sentence imposition. We build on his notion and conceptualize the consequences of sentencing two virtually identical cases in two subsections with different sentencing ranges and composition.

Upon crossing the quantifiable threshold, sentencers' consideration changes in two aspects. On the one hand, case B is sentenced according to a higher sentencing range. We refer to the difference between sentence B and sentence A caused by different sentencing ranges as the *severity* effect. On the other hand, case B is also compared to arguably more serious cases within its subsection and thus the relative position within the sentencing range is lower. We call the difference between sentence B and sentence A caused by different composition of cases within subsections as the *reference* effect.

To demonstrate the two mechanisms causing the severity and the reference effects, consider two policy changes that isolate the mechanisms. Suppose that sentencing ranges

[6]Sections 153 and 159.

change, but the composition of the offense subsection remains the same. Then only the severity mechanism is active. Keeping the initial rank of cases ordered by their relative seriousness, sentencers fit the sentences imposed within the new sentencing range. Thus, if a sentencing range for a specific offense subsection is increased (either the lower or the upper limit is increased), sentences are likely to increase as well.

However, if only the composition of cases within offense subsections changes – more or less severe cases are added or removed – positions of a specific case are transposed within the offense subsection. If the change in the composition points in one direction, the effect of the reference mechanism should be straightforward. If less serious cases than the current ones are added to an offense subsection, sentences for the current ones should increase since all current cases suddenly rank as relatively more serious within the offense subsection.

The severity and the reference mechanisms generally affect sentencing consideration in opposite directions. Depending on which effect dominates, the threshold either increases (B is sentenced to harsher sentence) or decreases (A is sentenced to harsher sentence) the sentence. The latter leads to perverse sentencing when (marginally) more serious cases are punished more leniently.

Literature studying the consequences of thresholds on sentencing is limited, with only several studies on the United States and on Russia. Using observable data from the United States, Bjerk (2017b) documents that judges imposed different sentences to offenders with a drug amount 10% above and below quantity thresholds. However, these effects became smaller or disappeared completely once he controlled for other observed characteristics such as weapons charge. Studying the threshold for 10 year mandatory minimums, Tuttle (2019) finds an important increase upon crossing the threshold. The effect was largely driven by prosecutorial discretion.

The effects of mandatory minimums cannot by easily generalized to lower limits of sentencing ranges. High mandatory minimums (5 and 10 years) and large prosecutorial discretion lead to a bunching of cases just below quantity thresholds (Bjerk, 2017b, 2005; Tuttle, 2019). Most criminal justice systems around the world have, however, less developed plea-bargaining system than the United States (Johnson, 2019); the prosecution has less power and, as a result, it cannot similarly influence sentencing outcomes around thresholds. Furthermore, mandatory minimums are often not applied even during sentencing. Specifically, mandatory minimums were imposed to less than half of drug offenders who were eligible for it at sentencing (United States Sentencing Commission,

2011, Chapter 8). Altogether, in other systems, prosecutors, offenders, and judges have fewer opportunities and less power to place the case above or below the threshold, to impose sentences below lower sentencing range, and they have less incentive to do so.

In a study most relevant to ours, Skugarevskiy (2017) examines sentencing for drug offenses in Russia, where he finds that crossing a threshold of 100 grams for cannabis or 2.5 grams for heroin leads to an increase of 0.84 years of imprisonment. His findings cannot be generalized to the effects of quantity thresholds and sentencing ranges, because in his setting sentencing ranges do not overlap. For offenses below the threshold, the sentencing range is 0-3 years of imprisonment, while above the threshold the sentencing range is 3-10 years. Such provisions imply that every offense over the quantity threshold warrants a sentence at least as serious as the sentence imposed for every individual offense below the threshold, ignoring a vast array of mitigating or aggravating factors. Such legislative provision introduces unwarranted disparities by definition, disabling the quantification of thresholds' role in systems with overlapping sentencing ranges of offense subsections.

2.2.2 Stylized Framework

We next introduce a framework that formalizes the previous discussion on the role of thresholds in sentencing decisions. Suppose that an offense can be fully characterized by two random variables with a joint probability distribution function $f_{X,T}(x,t)$, where X represents factors of the offense and T is a classifying variable (e.g. amount of drugs possessed).

A sentencing process is a two-stages rule whereby any offense (x,t) is assigned a sentence s. The rule is characterized by a set of thresholds $\tau \in \mathcal{T} = \{\tau^{[0]}, \tau^{[1]}, \tau^{[2]}, \ldots\}$ and corresponding sentencing ranges, i.e. intervals $\rho(\tau) = (\rho^{-}(\tau), \rho^{+}(\tau))$ that restrict the space for a possible sentence.[7] In the first stage, the rule classifies an offense (x,t) into a corresponding subsection $\tilde{\tau}$ by comparing the realization of the classifying variable t with the set of thresholds. In the second stage, the rule suggests a sentence s based on: (i) the sentencing range of the corresponding offense subsection $\rho(\tilde{\tau})$; and (ii) the relative seriousness of the particular case within the corresponding subsection measured by both factors x and the classifying variable t.

[7] Abusing the notation, we use τ as both the value of the upper limit of the classifying variable for a given subsection and as a label of that subsection itself.

Definition 1 (Sentencing Rule)**.** Suppose an offense (x, t). Then the sentence s assigned to this case is determined by the following two-steps *sentencing rule*:

$$\tilde{\tau} = min(\tau \in \mathcal{T} | \tau \geq t) \tag{D 1.1}$$

$$s = \rho^{-}(\tilde{\tau}) + G(x, t; q(\tilde{\tau}))(\rho^{+}(\tilde{\tau}) - \rho^{-}(\tilde{\tau})), \tag{D 1.2}$$

Function $G(x, t; q(\tau))$ – relative seriousness – determines the relative position of an offense (x, t) within a sentencing range of a subsection τ with a reference seriousness $q(\tau)$. The reference seriousness $q(\tau)$ captures the notion that the same offense is likely to be viewed as less serious when compared to a composition of more serious cases and *vice versa*. Since we assume that the recommended sentence must be within the corresponding sentencing range, function G is bounded between 0 and 1.

We further assume three properties of G. First, keeping everything else the same, as factors x of an offense increase, so does the offense's relative position within the sentencing range. In particular, we rely on a weak version of that property. Second, the same holds for a classifying variable t. Third, as the reference seriousness $q(\tau)$ increases (e.g. more serious cases are added to a subsection), the same offense is viewed as less serious and will be positioned lower in the sentencing range. Assumption 1 introduces the properties formally.

Assumption 1 (Relative seriousness)**.** *For any offense (x, t), any $\epsilon > 0$, and for any two levels of reference seriousness of q and q' such that $q < q'$, the following holds*

$$G(x, t; q(\tau)) \leq G(x + \epsilon, t; q(\tau)) \tag{A 1.1}$$

$$G(x, t; q(\tau)) \leq G(x, t + \epsilon; q(\tau)) \tag{A 1.2}$$

$$G(x, t; q') < G(x, t; q) \tag{A 1.3}$$

To avoid less intuitive and trivial cases, we assume that sentencing ranges are harsher in higher subsections. In particular, we assume that at least one of the limits of the sentencing ranges increases as we move to a more severe sentencing range. Since this seems to be true in all the criminal codes we know, this assumption is mostly technical and not controversial.

Assumption 2 (An Increase of Sentencing Ranges)**.** $\forall \tau, \tau' \in \mathcal{T} : \tau < \tau'$*, the following*

holds

$$\rho^-(\tau) \leq \rho^-(\tau') \tag{A 2.1}$$

$$\rho^+(\tau) \leq \rho^+(\tau') \tag{A 2.2}$$

$$(\rho^-(\tau') - \rho^-(\tau)) \times (\rho^+(\tau') - \rho^+(\tau)) > 0 \tag{A 2.3}$$

The aim is to use the introduced framework to study a situation in which the existence of a threshold causes two marginally different cases – cases that differ only in the classifying variable – to be sentenced according to different subsections. We next introduce a theorem that defines the problem formally and provides a solution to it.[8]

Theorem 1 (Difference in Sentencing)**.** *Suppose a set of thresholds $\mathcal{T}$ and two cases of the same offense (x,t) and $(x,t+\epsilon)$, where $\epsilon > 0$, but $\epsilon \to 0$, i.e. two cases that marginally differ in the value of the classifying variable t. Suppose further that $\exists \tau \in \mathcal{T}$ such that $t < \tau < t+\epsilon$. Then,*

$$\Delta s = \underbrace{\Delta\rho^- \left(1 - G(x,t;q(\tilde{\tau}_1)\right) + \Delta\rho^+ \left(G(x,t;q(\tilde{\tau}_1))\right)}_{\textit{severity effect}} + \underbrace{\Delta G \left(\rho^+(\tilde{\tau}_2) - \rho^-(\tilde{\tau}_2)\right)}_{\textit{reference effect}}, \tag{T 1.1}$$

where $\tilde{\tau}_1 = min\{\tau | \tau > t\}$ and $\tilde{\tau}_2 = min\{\tau | \tau > t+\epsilon\}$, $\Delta\rho^- = \rho^-(\tau_2) - \rho^-(\tau_1)$, $\Delta\rho^+ = \rho^+(\tau_2) - \rho^+(\tau_1)$, $\Delta G = G(x,t+\epsilon;q(\tilde{\tau}_2)) - G(x,t;q(\tilde{\tau}_1))$.

To sketch the proof, note that $\tilde{\tau}_2 > \tilde{\tau}_1$. Using (D1.2), the problem simplifies as follows[9]

$$\begin{aligned}\Delta s &= \rho^-(\tilde{\tau}_2) + G(x,t+\epsilon;q(\tilde{\tau}_2))(\rho^+(\tilde{\tau}_2) - \rho^-(\tilde{\tau}_2)) - \left(\rho^-(\tilde{\tau}_1) + G(x,t;q(\tilde{\tau}_1))(\rho^+(\tilde{\tau}_1) - \rho^-(\tilde{\tau}_1))\right) \\ &= \Delta\rho^- \left(1 - G(x,t;q(\tilde{\tau}_1))\right) + \Delta\rho^+ \left(G(x,t;q(\tilde{\tau}_1))\right) + \Delta G \left(\rho^+(\tilde{\tau}_2) - \rho^-(\tilde{\tau}_2)\right)\end{aligned}$$

Theorem 1 provides us with a clear prediction of the difference between two cases that marginally differ in the classifying variable around a threshold. The sign of the difference is, however, ambiguous, as there are two likely antagonistic effects: the *severity effect*, which is always non-negative[10] and the *reference effect*, which can be (and under likely occurring circumstances will be) negative. If the effects work in opposite directions, then

[8] The problem can be also formulated differently. Suppose an offense (x,t) and two different sentencing rules. Thresholds in the first rule are $\tau^{[1]}, \tau^{[2]}, \tau^{[3]}$; thresholds in the later rule are $\tau^{[1]}, \tau^{[2]} + \epsilon, \tau^{[3]}$, where $\tau^{[2]} < t < \tau^{[2]} + \epsilon$. Then the problem corresponds to the question of how the same offense (x,t) will be sentenced under two sentencing rules that marginally differ.

[9] See the Appendix for more details.

[10] In fact, the *severity effect* is likely to be positive.

the sign of the overall effect depends on the relative strength of these effects. We next discuss the signs of the two effects in more detail.

Corollary 1 (Sign of Severity Effect)**.** *The expression* $\Delta\rho^{-}\left(1-G(x,t;q(\tilde{\tau}_1))\right)+\Delta\rho^{+}\left(G(x,t;q(\tilde{\tau}_1))\right)$ *– the severity effect – is always non-negative.*

To see that the *severity effect* is always non-negative, note that assumption 2 implies that $\Delta\rho^{-}$ and $\Delta\rho^{+}$ are non-negative and since $G(x,t;q(\tau)) \in (0,1)$, the *severity effect* must be non-negative, too. Furthermore, the *severity effect* is zero only in two specific cases: (i) $\Delta\rho^{-}=0$ and $(G(x,t;q(\tilde{\tau}_1)))=0$, i.e. the lower limits of the sentencing ranges are the same and the case (x,t) is the least serious case in its subsection; and (ii) $\Delta\rho^{+}=0$ and $(G(x,t;q(\tilde{\tau}_1)))=1$, i.e. the upper limits of the sentencing ranges are the same and the case (x,t) is the most serious case in its subsection. Therefore, in what follows, we consider the *severity effect* positive.

Corollary 2 (Sign of Reference Effect)**.** *If* $q(\tilde{\tau}_1) \leq q(\tilde{\tau}_2)$*, then expression* $\Delta G\left(\rho^{+}(\tilde{\tau}_2)-\rho^{-}(\tilde{\tau}_2)\right)$ *– the reference effect – is negative.*

Note that the sign of the *reference effect* corresponds to the sign of ΔG, as any sentencing range $(\rho^{+}(\tilde{\tau})-\rho^{-}(\tilde{\tau}))$ is positive by definition. To determine the sign of $G(x,t;q(\tilde{\tau}_2))-G(x,t;q(\tilde{\tau}_1))$ note that assumption 1 implies that as long as $q(\tilde{\tau}_1) < q(\tilde{\tau}_2)$, i.e. the reference seriousness of the lower subsection is lower than the reference seriousness of the higher substitution, the reference effect is negative.

Implications

The introduced framework allows us to discuss how the structure of the thresholds shapes sentencing more generally. In some countries, such as in France, the sentencing ranges are organized so the lower limit of the sentencing range is the same across more (all) offense subsections and only the upper limit increases for subsections for higher values of classifying variables. Our framework helps us to understand how the final sentences will differ in this system compared to a system in which both the lower and the upper limits increase for subsections for higher values of classifying variables.

Let us denote the two systems as α and β and assume an offense that has two subsections. The thresholds are identical in both systems ($\mathcal{T}^{\alpha}=\mathcal{T}^{\beta}$), Furthermore, the upper limits of sentencing ranges are the same in both systems, $\rho^{+}_{\alpha}(\tau)=\rho^{+}_{\beta}(\tau)\ \forall\tau$. The two systems differ only in the lower limit of the sentencing ranges. In particular, in the

Figure 2.2: Two different structures of sentencing ranges

Notes: This figure graphically compares two structures of sentencing ranges in subsections. The left panel represents a system in which the lower limits of the sentencing ranges are the same in all subsections (French-like system), while the right panel represents a system in which both the lower and the upper limits increase in a *higher* subsection.

α system, the lower limits of sentencing ranges remain the same for both subsections $\rho_\alpha^-(\tau^0) = \rho_\alpha^-(\tau^1)$, while in the β system, the sentencing range is increasing. Additionally, we assume that for τ^0 both systems have the same lower limits of the sentencing range. Therefore, $\rho_\alpha^-(\tau^O) = \rho_\alpha^-(\tau^1) = \rho_\beta^-(\tau^0) < \rho_\beta^-(\tau^1)$. Figure 2.2 represents the structure of both systems. Note that neither of the systems violates our Assumption 2.

Our framework implies that the severity effect is weaker in the α (French-like) system. Consequently, the expected difference between two marginally different sentencing decisions around the threshold is lower and the probability of a perverse sentencing is higher. This follows from Theorem 1 and the fact that $\Delta\rho_\alpha^- = 0 < \Delta\rho_\beta^-$. This comparative statics assumes that the $G(x, t; q(\tau))$ is the same in both systems.

Whether and how fast the lower limits of sentencing ranges grow with a higher subsection of an offense should reflect policy-makers' concerns regarding the type of injustice the sentencing system can introduce. If a policy-maker is more concerned that the structure of the criminal code would cause perverse sentencing, i.e. a less severe case will be sentenced to longer incarceration, then the lower limits should be increasing with a higher subsection. Conversely, if the concern is that a slightly more severe case will be sentenced to an unreasonably harsher punishment, then the lower limit should remain the same (or grow slower) across different subsections.

2.3 Experiment with Prosecutors

The aim of our experiment is twofold. First, we quantify the consequences of thresholds on the average recommended sentence in a controlled environment with professional sentencers. Despite the insights of the theoretical model, the sign and the magnitude of the effect is an empirical question. Second, we design an experimental treatment to test for

the existence of the severity and the reference mechanisms.

2.3.1 Background

The Czech criminal justice system is a typical continental European legal system similar to the German one. The Criminal Code and Code of Criminal Procedure establish rules with the higher courts' jurisprudence playing only a quasi-precedential role, as precedents are not binding. The Criminal Code defines both offenses and sentencing ranges for either the entire offense or for its subsections. As there are no sentencing guidelines, judges have wide discretion as to what type of sanction and of what quantity to impose. Plea bargains are rare, resulting in judges imposing sentences in virtually all criminal cases.

The prosecution's role in sentencing consists of recommending sanctions and appealing against a sentence. Until 2019, prosecutors would recommend sanctions either only in the closing argument or also in the indictment. As of January 2020, (six months prior to our experiment) they are obliged to recommend a specific sanction along with the indictment. Prosecutors can file an appeal arguing that either an error of law or fact was made; no specific level of an error (i.e. substantial or palpable) is required. Judges are highly incentivized to appease both the defendant and the prosecutor so that both sides waive their rights to an appeal. Should there be no appeals, judges can issue a simplified judgment and do not have to provide detailed written reasoning. Prosecutors are aware of their power over the imposed sentences. Drápal and Dušek (2021) document that the efforts of Czech prosecutors in 2016 led to a high increase in the fine imposition rate.

Regarding mentality, Czech prosecutors are not law-and-order punitive players. This was documented by the Prosecutor General, who suggested lowering sentencing ranges in order to reduce the high prison population in the Czech Republic (Zeman, 2020). Regarding organizational structure, while the Prosecutor General can issue binding orders in general matters (such as regarding recommendation of a specific sentence), it generally cannot interfere with individual cases. This is also due to the fourth-level hierarchical structure of prosecution. District offices prosecute the vast majority of cases (98%), while regional and high offices deal with the most serious cases. The Prosecutor General's Office files appeals in legal matters to the Supreme Court and unifies the practices of lower offices. The head of a higher prosecution office can only directly influence the cases handled by a prosecution office of one lower level, rendering regional prosecution offices the most important for influencing every-day practice. Heads of each prosecution office

influence sentence recommendations directly as they approve all indictments and recommended sentences. The Prosecutor General thus has only a limited role in influencing the sentencing practices of first-level prosecutors.

Prosecutors are appointed by the Minister of Justice following a proposal by the Prosecutor General after serving at least 3 years as assistants to a prosecutor, to a judge or to an attorney and after passing a professional exam. Prosecutors have tenure and they can be removed only via disciplinary proceedings if they commit blatant mistakes. While we focus on sentencing, prosecutors' duties include overseeing investigations, filing indictments, and participating in court hearings. Anecdotal evidence suggests that prosecutors pay more attention to the guilt decision than to the one on sentencing.

2.3.2 Experimental Design

The theoretical framework helps us to design scenarios of criminal cases with varying parameters so the results are informative about the size and magnitude of the effect caused by the thresholds and existence of the two proposed mechanisms. To estimate the overall effect, it is sufficient to rely on the existing legislation and vary the amount of the classifying variable so that two presented cases differ only marginally but are sentenced according to different offense subsections. A comparison between the average recommended sentences then yields causal estimates of the effect of thresholds on sentencing.

Theorem 1 implies that testing either for the reference or the severity effects requires that the other mechanism be muted. To test for the severity effect, the reference effect must be suppressed and *vice versa*. To attain such comparison, we introduce a hypothetical legal framework that varies the sentencing ranges and offense compositions. In particular, to test for the reference effect, one needs to compare scenarios in which two identical cases are sentenced according to subsections with the same sentencing ranges, but with a different structure of cases. We generate the variation by adjusting the thresholds and thus effectively adding more severe criminal cases into the subsection. Similarly, to test for the severity effect, we compare two scenarios for which the relative position among other cases remains the same, but the upper limits of the sentencing ranges differ. To minimize inconsistency in the criminal code in these hypothetical situations, we rely only on small changes in sentencing ranges. Such changes increase the credibility of the hypothetical scenario, as it resembles the existing legislation known to prosecutors, but it also suppresses the predicted effect. Tests for the reference and the severity effects are

implemented in a drug possession case.[11]

2.3.3 Implementation of Experimental Design

To approach prosecutors, we partnered with the Prosecutor General's Office of the Czech Republic. The invitation to participate in the survey was sent by a deputy to the Prosecutor General. Of the 1257 prosecutors invited to participate, 206 prosecutors responded to the first vignette and 194 to both of them. Each participant received a unique link, enabling us to eliminate duplicate answers from the same prosecutors. Additionally, the unique link allows us to match the experimental data with anonymized administrative data containing basic information about prosecutors such as their gender, age, time on the bench, position in a prosecutors' organization, alma mater, and whether they were members of the communist party up to the 1989 Velvet Revolution.[12] The relatively low participation rate of less than 20 % limits our power to identify statistically small effects, but does not harm our interpretation of the results.

Participants are not a representative sample of the population of prosecutors. Compared with the population of prosecutors, our sample contains a higher proportion of female respondents, those who work at the Prosecutor General's Office, and the participants are, on average, younger. It is likely that the overrepresentation of prosecutors from the Prosecutor General's Office is caused by higher willingness to comply with the request of their direct superior.

The observed selection into participation may limit generalization of our results. If the selection interacts with the size of treatment effects e.g., female prosecutors tend to base recommended sentences more on the other factors than on the exact amount of the classifying variable, then the average treatment effect in a population of prosecutors may systematically differ from the estimated effect. Unfortunately, previous literature offers little or no evidence on heterogeneity in factors affecting sentencing decisions among prosecutors with different demographic characteristics. Importantly for our results, different treatment groups seem to be balanced in terms of observable characteristics and the treatment effect is thus unlikely to be driven by other factors. The descriptive statistics and balance tests are presented in Table 2.6 in the Appendix.

Each participant in the experiment was asked to analyse and recommend a length of

[11] The experiment was pre-registered as AEARCTR-0006023.

[12] We managed to match almost every experimental observation to administrative counterparts. Only a few observations could not be matched.

incarceration in two hypothetical criminal cases: (i) theft and (ii) drug possession.[13] Each case was described on one page, and the wording was consulted with several practising sentencers so as not to omit relevant informant. We also provided the respondents with the relevant section of the Criminal Code (the offense with multiple subsections) along with an excerpt from jurisprudence establishing the quantity thresholds (in the case of drug offense). The participants were warned on the introductory screen that the provisions may not be in line with those in force; this comment was emphasized in italics. After the participants recommended a length of incarceration, we asked them to indicate the offense subsection according to which they impose the sentence. Since the classification of the subsection is based on quantifiable variables that were provided in the case, there is an objectively right answer. Failure to identify the correct subsections can indicate inattentive or careless responses.

In the experiment, participants were presented with an order from the head of their prosecution office to impose a non-suspended prison sentence, while its length was left entirely for their consideration. Both cases were prepared such that the imposition of a non-suspended prison sentence would be the most likely choice for most prosecutors due to the extensive criminal history of defendants. To select the length of imprisonment, participants were supposed to choose both the number of years and months of imprisonment from a drop-down menu. Their answers were not limited to the relevant length of incarceration according to the Criminal Code, and thus it was technically possible to recommend longer or shorter sentences than those given by the Criminal Code.

Each criminal case was presented in four different scenarios (i.e. four treatment arms in both cases) that differ in parameters of the cases and generate the intended experimental variation. Upon entering the platform, participants were randomized into one of the treatment arms to solve one of the four different scenarios in the first case (drug possession). Once they submitted their recommended sentence in the first case, they were once again randomized in one of the scenarios in the second case. The randomization into a treatment arm in the second case (theft) was independent from the first randomization. Note that since not everyone who started the survey also finished it; the randomization based on entry does not necessarily lead to a numerically identical number of observations in each scenario. In a between-subjects design we explore the variation in the length of incarceration recommended in different treatment arms.

[13]The cases were presented in the reverse order to everyone.

2.3.4 Theft

Vignette description

Prosecutors were asked to recommend a length of incarceration in the following case. An offender visited his parents to help them with their computer. Once he was left alone in their home, he took advantage of his parents being logged into their internet banking account, and of their cellphone being left at home, and transferred all their money to his own account. He gambled away all of the money in the following three days. He had previously been sentenced for fraud and embezzlement and he had been released from a 2 year prison sentence two months prior to committing the offense in the vignette.

The full text of the vignette, along with the text of the section and other information provided to the participants, is presented both in English and in Czech in the Appendix. We randomized participants into four treatment arms (scenarios), as captured by Table 2.1. In the case of theft, we did not include scenarios that vary the existing legislation. Instead, we study overall threshold effects around two sequential 50k and 500k thresholds.[14]

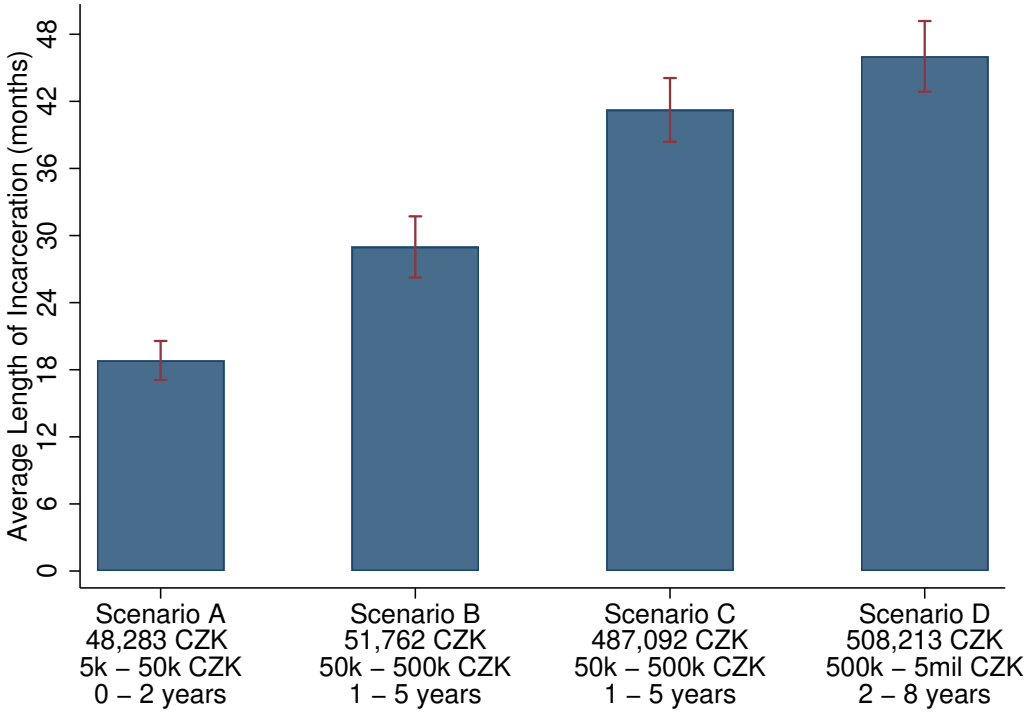
Table 2.1: Scenarios of Theft Offenses

Scenario	Damage Caused (CZK)	Subsection Composition (CZK)	Sentencing Range (years of incarceration)
A	48 283	5 000 – 50 000	0 – 2
B	51 283	50 000 – 500 000	1 – 5
C	487 092	50 000 – 500 000	1 – 5
D	508 213	500 000 – 5 000 000	2 – 8

We estimate three causal effects regarding the theft case. First, a comparison of the average length of incarceration recommended in scenarios A and B yields the effect of the 50k threshold. Second, a comparison between scenarios C and D estimates the change in the length of incarceration caused by the 500k threshold. Third, to investigate what the effect of a dramatic increase in the classifying variable is, we use the fact that scenario B and C differ only in the size of the caused damage, but not in the subsection and sentencing ranges. Therefore, the comparison between the average recommended sentence estimates the causal effect of a dramatic increase in damage caused.

[14]Several months after our experiment, in October 2020, the thresholds doubled. There is also an additional subsection for repeated thefts capped by the 50k threshold with a higher sentencing range (0.5-3 years) which is not included in our experiment.

Figure 2.3: Theft: Average Recommended Sentence by Prosecutors



Notes: The graph shows the average length of incarceration by different scenarios. The average sentence in scenario A is 18.8 months, in scenario B 29 months, in scenario C 41.2 months, and in scenario D 46 months. 95% confidence intervals are displayed.

Finally, we are also interested in a comparison of the 50k threshold effect and the effect caused by increasing the damage almost 10 times. Formally, we estimate the following $(s(C) - s(B)) - (s(B) - s(A))$ and test whether it equals to zero. The estimate provides us with a meaningful and contextual interpretation of the magnitude of the effect caused by the 50k threshold in terms of the magnitude of the classifying variable.

Results

Figure 2.3 shows the average length of incarceration recommended by prosecutors in different scenarios. The higher the damage caused, the longer the recommended sentence. In scenario A, the average length of incarceration is slightly above 18 months. In scenario B, the average sentence is higher by 10 months. In the remaining scenarios C and D, the average sentences are approximately 41 and 46 months, respectively.

To test the effects formally, we estimate three models. Model 1 represents a simple univariate OLS regression in which the treatment effect is captured by a dummy that

equals to 1 if the observation comes from the treated scenario and 0 otherwise. For the formal regression, see equation Model 1. Model 2 extends the univariate OLS by controlling for additional characteristics of prosecutors (age, age^2, $gender$, $Hierarchy$ - position in the system of prosecutors). Finally, Model 3 estimates a univariate OLS on a sample including a number of participants who did not identify the correct subsection.

$$Sentence = \alpha + \beta Treatment + \epsilon \tag{Model 1}$$

$$Sentence = \alpha + \beta Treatment + \delta_1 Male + \delta_2 Age + \delta_3 Age^2 + \sum_{i=1}^{4} \gamma_i Hierarchy_i + \epsilon \tag{Model 2}$$

Table 2.2 shows three panels, each devoted to one of the three questions: the effect of the 50k threshold, the effect of the 500k threshold, and the effect of size of the classifying variable. Panel I presents robust evidence that the 50k threshold increases the average incarceration by about 10 months, which represents more than a 50% increase compared to the sentence for cases just below the 50k threshold. Once we control for gender, age, and position in the hierarchy of prosecutors, the point estimate is even larger. Panel II estimates the effect of the 500k threshold. Both Model 1 and Model 2 suggest that the 500k threshold increases the average length of incarceration by 4.5 months (approx. 10%). While the effect of the 500k threshold is smaller than the one caused by the 50k threshold, it provides additional evidence that the existence of the effect is rather general and limits the external validity concerns. Finally, Panel III presents evidence that increasing the damage almost 10 times increases the sentence by roughly a year.

The effect of the size of damage allows us to understand the enormous effect the 50k threshold caused. Increasing a damage by CZK 435,000 (by 855%) corresponds to an additional 12 months of incarceration. This contrasts to the 10 month increase in incarceration for a theft that causes higher damage by CZK 3,000 estimated in Panel I. Alternatively, we can compare the 50k threshold effect to the effect caused by increasing the damage by CZK 435,000. Formally, we run the following regression

$$\text{Sentence} = \beta_1 \text{Scenario A} + \beta_2 \text{Scenario B} + \beta_3 \text{Scenario C} + \varepsilon \tag{2.1}$$

Table 2.2: The Effect of Quantity Thresholds on Sentence Decisions for Theft

Panel I: 50k Threshold

	Model 1	Model 2	Model 3
Treatment Effect	10.153***	11.631***	10.056***
	(1.614)	(1.683)	(1.595)
Constant	18.826***	11.730**	18.826***
	(0.866)	(5.402)	(0.866)
Control Variables	No	Yes	No
Only if Correct Subsection	Yes	Yes	No
N	96	92	97

Panel II: 500k Threshold

	Model 1	Model 2	Model 3
Treatment Effect	4.789**	4.436**	3.047
	(2.110)	(2.200)	(2.272)
Constant	41.235***	22.847	41.235***
	(1.420)	(22.947)	(1.420)
Control Variables	No	Yes	No
Only if Correct Subsection	Yes	Yes	No
N	92	85	97

Panel III: Size of Damage

	Model 1	Model 2	Model 3
Treatment Effect	12.255***	11.705***	12.353***
	(1.966)	(2.073)	(1.950)
Constant	28.98***	38.615***	28.882***
	(1.361)	(7.419)	(1.338)
Control Variables	No	Yes	No
Only if Correct Subsection	Yes	Yes	No
N	101	97	102

Robust standard errors in parentheses.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Panel I tests the 50k threshold effect (s(B) - s(A)), Panel II tests the 500k threshold effect (s(D) - s(C)), and Panel III tests the effect of the size of damage (s(C) - s(B)). *Treatment* corresponds to a dummy variable which equals 1 for respondents who were assigned to the treatment group with a higher damage.

Model 1 represents a simple univariate OLS regression, Model 2 extends the univariate OLS by controlling for additional characteristics (*age*, age^2, gender, position in the system of state prosecutors), Model 3 a univariate OLS that includes responses of participants who did not identify the correct subsection of the paragraph.

and using the Wald test, we test

$$\beta_1 - 2\beta_2 + \beta_3 = 0$$

We do not reject the null hypothesis suggesting that the 50k threshold has the same effect on the absolute length of incarceration as increasing the damage by CZK 435,000 (by 855%). Despite the probable sensitivity of the results on particular specifications and circumstances of cases, we take the results as robust evidence that the effect of thresholds on the shape of sentencing decisions is of great importance.

2.3.5 Drug Possession

Vignette description

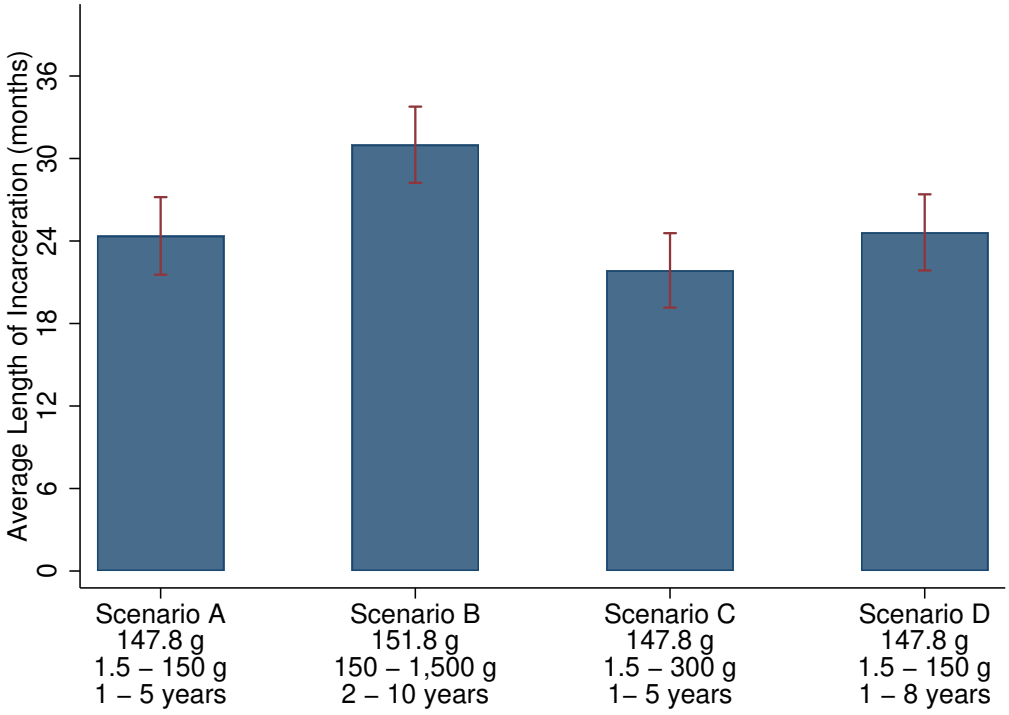
In the drug possession case, an offender was selling methamphetamine in front of a dance club in a town in Northern Bohemia. It was a one-time event after the offender lost his job. He had been repeatedly sentenced for selling marijuana over the previous 10 years and for small thefts. In four scenarios, we manipulate: (i) the amount of pure substance of methamphetamine (henceforth only methamphetamine) found on him; (ii) the applicable sentencing range; and (iii) the thresholds influencing the composition of cases within corresponding subsections. The full text of the vignette, along with the text of the section and other information provided to the participants, is presented both in English and in Czech in the Appendix. The four scenarios that were applied are summarized in Table 2.3. Scenario A and scenario B are based on the existing legislation and differ marginally in the amount of the drug possessed. The difference between the length of incarceration recommended in scenario A and B represents the overall threshold effect caused by the threshold.

Table 2.3: Scenarios of Drug Possession Offenses

Scenario	Amount Possessed (grams)	Subsection Composition (grams)	Sentencing Range (years of incarceration)
A	147.8	1.5 – 150	1 – 5
B	151.8	150 – 1500	2 – 10
C	147.8	1.5 – 300	1 – 5
D	147.8	1.5 – 150	1 – 8

Scenarios C and D introduce an alternative legal framework. In particular, in sce-

Figure 2.4: Drug Possession: Average Recommended Sentence by Prosecutors



Notes: The average sentence in scenario A is 24.4 months, in scenario B 31 months, in scenario B 21.9 months, and in scenario D 24.6 months. 95% confidence intervals are displayed.

nario C, we move the threshold of the classifying variable from 150 g to 300 g of methamphetamine and thus effectively add relatively more severe cases in the relevant subsection. It follows that a comparison between A and C scenarios captures the reference mechanism. In scenario D, we move the upper limit of the sentencing range and thus evoke the severity mechanism. Note, however, that the magnitude of the reference and severity mechanisms estimated here are only fractions of the whole reference and severity effect that work in the overall threshold effect.

Results

Figure 2.4 shows the average sentence recommended by prosecutors in different scenarios. In scenario A, the average length of incarceration is slightly more than 2 years and in scenario B longer by more than 6 months. The figure also reveals that the lowest average sentence is in scenario C, slightly more than 22 months, and finally, scenario D is almost identical to scenario A.

Table 2.4 shows results from formal tests of three hypotheses; Panel I tests the overall

threshold effect, Panel II tests for the reference mechanism, and Panel III tests for the severity mechanism. Similarly to the first case, each panel shows results from three empirical models: (i) a simple univariate OLS regression – conveniently, in all three panels the treatment dummy equals to zero for observations from scenario A; (ii) an OLS model controlling for additional characteristics of prosecutors – for the formal regression see equation Model 1; and (iii) a univariate OLS on a sample including several participants who did not identify the correct subsection.

Panel I of Table 2.4 provides convincing evidence that the threshold increases the length of incarceration by around 6 months, which represents approximately a 25% increase of the length. Employing different models, the point estimates of the treatment effect range from 5.2 to 6.6 months. The results thus provide robust evidence that prosecutors recommended a significantly different length of incarceration for two almost identical cases (the difference is 3 grams of methamphetamine (2.7%)) .

Panel II of Table 2.4 shows suggestive evidence of the reference mechanism. All three specifications provide similar negative point estimates, suggesting that by increasing the threshold of the classifying variable from 150g to 300g, the sentence decreases by 2.5 - 3 months. The effect represents approximately a 10% decrease in the length of incarceration. Since only one point estimate is statistically significant, we consider the results only as suggestive evidence.

Finally, three specifications in Panel III test for the severity mechanism and suggest the null results. To understand why the severity mechanism is rather negligible, it is important to realize that it does not test the full severity effect, but rather the existence of the mechanism. The average sentence in scenario A is located around a fourth of the possible interval (a year above the minimum of the sentencing range and 3 years below the maximum of the sentencing range). Thus, increasing the upper limit of the sentencing range further has only a limited impact.[15] Importantly, since the overall threshold effect is sizeable and positive, our framework implies that the full severity effect must also be sizeable and positive.

[15]This may be deemed a shortcoming of the experimental design. However, at the time of preparing the experimental design, we did not know what the average sentence in scenario A would be.

Table 2.4: The Effect of Quantity Thresholds on Sentence Decisions for Drug Possession

Panel I: Effect of Threshold

	Model 1	Model 2	Model 3
Treatment Effect	6.629***	5.870***	5.240**
	(1.969)	(2.229)	(2.016)
Constant	24.370***	41.876	25.379***
	(1.407)	(25.440)	(1.449)
Control Variables	No	Yes	No
Only if Correct Subsection	Yes	Yes	No
N	103	98	108

Panel II: Reference Effect

	Model 1	Model 2	Model 3
Treatment Effect	-2.506	-2.573	-3.516*
	(1.944)	(1.974)	(2.003)
Constant	25.379***	42.851	25.379***
	(1.407)	(26.304)	(1.449)
Control Variables	No	Yes	No
Only if Correct Subsection	Yes	Yes	No
N	98	96	102

Panel III: Severity Effect

	Model 1	Model 2	Model 3
Treatment Effect	0.257	-0.808	-0.009
	(1.969)	(2.095)	(1.993)
Constant	24.370***	40.990***	25.379***
	(1.407)	(8.204)	(1.448)
Control Variables	No	Yes	No
Only if Correct Subsection	Yes	Yes	No
N	105	102	112

Robust standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Panel I tests the overall threshold effect (s(B) - s(A)); Panel II tests the reference effect (s(C) - s(A)); and Panel III tests the severity effect (s(D) - s(A)). *Treatment* corresponds to a dummy variable which equals 0 for respondents who were assigned to the A scenario and 1 otherwise.

Model 1 represents a simple univariate OLS regression, Model 2 extends the univariate OLS by controlling for additional characteristics (*age*, age^2, gender, position in the system of state prosecutors), Model 3 a univariate OLS that includes responses of participants who did not identify the correct subsection of the paragraph.

2.4 Just Sentence and Its Measure

To provide insights beyond differences in the average recommended sentences, we propose a measure of justice that quantifies the probability that an observed sentence is just. Suppose two cases and sentences imposed for them. Two forms of injustice may arise. On the one hand, a more lenient sentence might be imposed for a more serious case. On the other hand, even if a harsher sentence is imposed for a more serious offense, such sentence can still be considered unjust if it is unreasonably harsher. Therefore, a necessary requirement to consider a sentence *just* is that the more serious case leads to a harsher sentence, but not excessively so. To avoid a normative stance and using a specific measure on what is considered an unreasonably harsher sentence, we measure the shares of sentences that are not considered just conditionally on different levels of harshness.

We first define the *just* sentence formally. Using the notation introduced earlier, consider a case (x, t) with a corresponding sentence s_1. Given this sentence s_1, we ask whether a different sentence s_2 in a potentially more severe case $(x + \delta_x, t + \delta_t)$, with $\delta_x \geq 0$ and $\delta_t \geq 0$ is *just* or not. We say that the sentence s_2 is *just* if it is: (i) at least as harsh as s_1; and (ii) is not unreasonably harsher. In other words, we take a random pair of cases and corresponding sentences and compare whether and by how much the sentence in a more serious case is harsher. The sentence in a more serious case is *just* if it is reasonably harsher. To enable different perceptions of unreasonable harshness, the definition is parametric.

Definition 2 (Just Sentence)**.** Given a sentence s_1 in a case (x, t) and a tolerance parameter $\eta \geq 1$, we say that a sentence s_2 of a case $(x+\delta_x, t+\delta_t)$, where $\delta_x \geq 0$ and $\delta_t \geq 0$ is *just* if the two following conditions are satisfied:

$$s_1 \leq s_2 \tag{D 2.1}$$

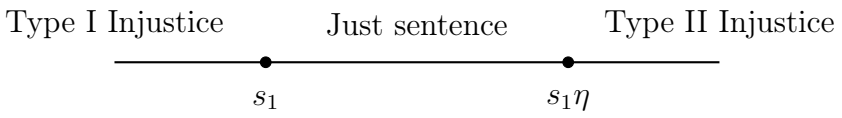
$$s_2 \leq s_1\eta. \tag{D 2.2}$$

The tolerance parameter η captures what is considered a reasonably harsher sentence and what is not. Importantly, the definition says that a sentence s_2 can be called *just* only if there is a tolerance parameter η and a sentence s_1 to which s_2 is compared. In other words, the just sentence is viewed relative to another sentence and a tolerance parameter. Without these, the definition is meaningless.

Definition 2 implies that a sentence that is not considered *just* can be of two types.

Depending on which condition is not satisfied, we distinguish two types of *unjust* sentences. If condition D 2.1 is not satisfied and the more severe case leads to a more lenient sentence, then we refer to *Type I injustice*, whereas if condition D 2.2 is not satisfied and the sentence s_2 is too harsh, we refer to *Type II injustice*. Figure 2.5 graphically represents the idea.

Figure 2.5: Just Sentence



Notes: Given a tolerance parameter η and sentencing decision s_1, then depending on the position of s_2, we define a *just* sentence, *Type I injustice*, or *Type II injustice*.

Having defined a *just* sentence and types of injustice, we next introduce empirical measures. To fit our experimental design, we consider the following situation. There are N^A sentencing decisions in a case $A = (x, t)$ denoted as s_i^A (e.g. a theft case with damage of CZK 48,283). There are also N^B decisions in cases $B = (x + \delta_x, t + \delta_t)$ denoted as s_j^B (e.g. a theft case with damage of CZK 51,283).

We introduce three empirical measures based on a similar logic. We compare each sentence s_i^A to each sentence s_j^B (i.e. $N^A \times N^B$ comparisons) a quantify frequency of $\frac{s_j}{s_i}$ being less than 1, between 1 and η, and higher than η. The obtained figures correspond to the measure of justice $\mathcal{M}^J(\eta)$, the measure of *Type I injustice* $\mathcal{M}^I$, and *Type II injustice* $\mathcal{M}^{II}(\eta)$. Note that *Type I injustice* does not depend on the tolerance parameter η. Intuitively, our measures quantify the probability that for a randomly observed pair of sentences s_i^A and s_j^B, the latter will be less severe, reasonably harsher, or unreasonably harsher compared to the former. The following definition introduces all three measures formally.

Definition 3 (Measures of Justice and Injustice)**.** Suppose that there are N^A decisions in a case $A = (x, t)$ denoted s_i^A and N^B decisions in cases $B = (x + \delta_x, t + \delta_t)$ denoted as s_j^B, in which $\delta_x \geq 0, \delta_t \geq 0$. Then for a given parameter η, the measure of *Justice* $\mathcal{M}^J(\eta)$, measure of *Type I injustice* $\mathcal{M}^I(\eta)$, and measure of *Type II injustice* $\mathcal{M}^{II}(\eta)$ are defined as follows

$$\mathcal{M}^J(\eta) = \frac{\sum_{i \in I} \sum_{j \in J} \mathbf{1}\left[1 \leq \frac{s_j^B}{s_i^A} \leq \eta\right]}{N^A \times N^B} \qquad \text{(Measure of } \textit{Justice}\text{)}$$

$$\mathcal{M}^I = \frac{\sum_{i\in I}\sum_{j\in J}\mathbf{1}\left[\frac{s_j^B}{s_i^A} < 1\right]}{N^A \times N^B} \qquad \text{(Measure of } \textit{Type I Injustice}\text{)}$$

$$\mathcal{M}^{II}(\eta) = \frac{\sum_{i\in I}\sum_{j\in J}\mathbf{1}\left[\eta < \frac{s_j^B}{s_i^A}\right]}{N^A \times N^B}. \qquad \text{(Measure of } \textit{Type II Injustice}\text{)}$$

The proposed measures have two properties. First, the sum of the measures equals to 1 for any given tolerance parameter η. Second, the higher the tolerance parameter η, the higher the *measure of justice*. Intuitively, as we increase the tolerance parameter, even harsher sentences are considered just (or tolerated). Naturally, by increasing the tolerance parameter, the measure of *Type II injustice* decreases.

2.4.1 Measure of Justice in the Experiment

We calculate the introduced measures for sentences recommended by prosecutors in our experiment. In particular, we focus on the threshold effect in the drug possession case, the 50k and the 500k thresholds effects in the theft case, and the size effect in the theft case. For each of these comparisons, we calculate *Type I injustice* and two measures of justice with tolerance parameters of 1.5 and 2. Finally, we report an approximation of how tolerant the society must be to evaluate at least half of the sentences as *just*.

Table 2.5: Measures of (In)Justice in the Experiment

	Type I injustice	$\mathcal{M}^J(1.5)$	$\mathcal{M}^J(2)$	$\left(\mathcal{M}^J\right)^{-1}(0.5)$
Drug Possession	0.21	0.41	0.62	1.75
Theft (50k threshold)	0.15	0.38	0.62	1.85
Theft (500k threshold)	0.33	0.49	0.63	1.55
Size of Damage	0.15	0.41	0.65	1.70

Table 2.5 shows the results. The *Type I injustice* is highest in the 500k threshold comparison, as a third of the more severe cases were sentenced to a more lenient sentence. This suggests that these two cases (theft cases with damage of CZK 487,092 and CZK 508,213) are perceived the most similar among all four pairwise comparisons. The second column implies that if the society tolerates that a more severe case is sentenced to a 50% longer incarceration, then the share of *just* sentences is around 40% for most of the comparisons, with an exception for the 500k threshold. Once the tolerance parameter is two, roughly two thirds of cases are considered fair in all four comparisons. Note, however, that the interpretation differs. For example the 50k threshold effect and the

500k threshold effects differ dramatically in the measure of *Type I injustice*, and since the measures sum to 1 for any η, they also differ in the *Type II injustice*.

Finally, an inverse function of *measure of justice* evaluated at 0.5 indicates how tolerant one has to be (i.e. what the η must be) to reach 50% probability that, from a randomly observed pair of sentences s_i^A and s_j^B, the latter is considered a just sentence. For example, the first row that compares a drug possession case with 148.8 grams and a drug possession case with 151.8 grams reveals that to reach 50% of just decisions, one has to consider a 75% longer incarceration for 3 grams of methamphetamine as a just sentence. This is additional evidence that the threshold has an enormous effect not only on the average sentence recommended. Figure 2.6 shows measure of justice and injustice for four main comparisons.

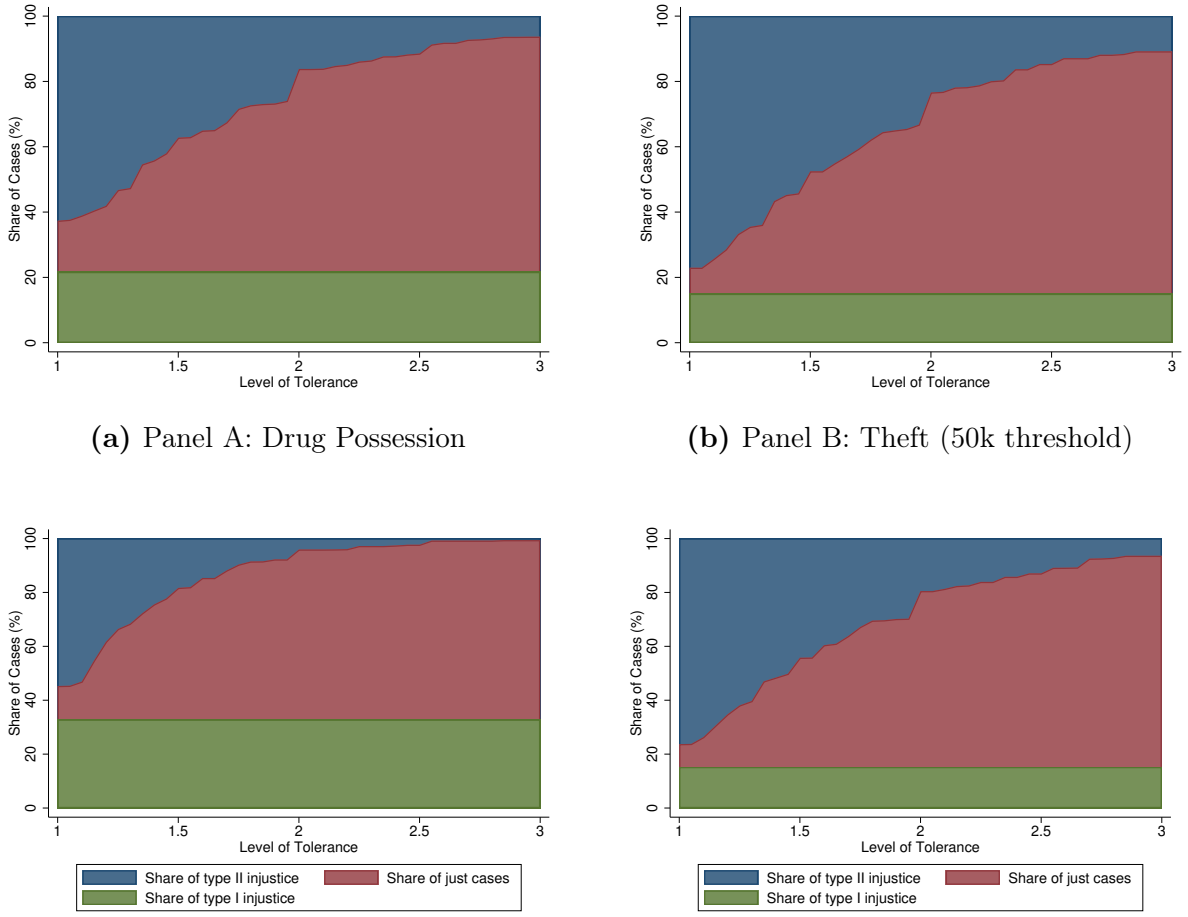
2.5 Discussion

Introducing a measure providing guidance to sentencers necessarily includes a trade-off. Measures envisioned to reduce sentencing disparities will likely introduce a certain amount of new ones. As we show, the extent of new disparities may be substantial and the disparities introduced by dividing offenses into subsections with specific sentencing ranges may strongly limit the benefits of such a measure. The difference in sentences imposed for identical thefts around a threshold (CZK 48,283 and CZK 51,283 scenarios) was as large as the difference in sentences imposed for identical thefts at the opposite sides of one offense subsection (CZK 51,283 and CZK 487,092). In other words, stealing more by CZK 3,000 leads to a similar increase in sentence as stealing more by CZK 435,000. Any criminal justice system striving for principled sentencing cannot tolerate such a disparity.

We develop a theory explaining how thresholds influence sentencing around them and what a change in the sentencing ranges and offense subsection composition can cause. The theory implies that a difference between two sentences can be decomposed into two opposing effects. The severity effect captures the difference if only the sentencing range changes and the composition of cases in a subsection remains the same. Conversely, if the sentencing range remains unchanged, but the subsection composition changes, we refer to the reference effect. In line with our theory, each of these two mechanisms is expected to influence all cases within a subsection in the same direction.

We empirically test for both mechanisms proposed. First, testing for the reference

Figure 2.6: Measures of (In)Justice in Experiment



(a) Panel A: Drug Possession

(b) Panel B: Theft (50k threshold)

(c) Panel C: Theft (500k threshold)

(d) Panel D: Theft (Size of Damage)

Notes: Panel A shows a measure of justice for the case of drug possession (scenario A vs. B). Panel B shows a measure of justice for the case of theft around the 50k threshold. Panel C shows a measure of justice for the case of theft around the 500k threshold. Panel D shows a measure of justice for the case of theft in terms of size of damage.

effect directly, we find suggestive evidence of a negative effect as predicted. Second, while the direct test for the severity effect provides the null results, the overall effect and the decomposition guided by our theory suggest that the the severity effect must dominate. The reason why, despite its arguable size, we fail to find evidence supporting the mechanism likely lies in the experimental design. The experimental design does not test for the full severity effect; instead it tests for a mechanism that corresponds to a part of the effect. To see the difference, note that the severity effect – as part of the decomposition of the threshold effect – consists of changing both the lower and the upper limits of the sentencing range dramatically, while the implemented variation changes only the upper limit and only marginally. Additionally, the effect is weakened by the fact that the initial average sentence is closer to the lower limit of the sentencing ranges which further attenuate the effect of an increase in the upper limit of the sentencing range.

More nuanced predictions would require additional assumptions on the sentencers' behavior and in particular on how sentencers fit sentences within ranges (i.e. on function $G(.)$). Still, based on our theory, we can make three observations about the expected size of the effects. First, consider the question of whether increasing the lower or the upper limit of the sentencing range will lead to a higher effect. The severity effect is a linear combination of an increase in the lower limit and the increase in the upper limit where the weights are determined by a relative position of the case within the sentencing range, i.e. $G(.)$. Therefore, sentences in more serious cases ($G(.) > 0.5$) are more sensitive to increases in the upper limit, while less serious cases ($G(.) < 0.5$) to increases in the lower limits. Since most of the sentences seem to be clustered in the lower part of the sentencing ranges close to the lower limit (Drápal, 2020; Sutela, 2020), increasing the lower limit would lead to a higher increase in the average sentence.

Second, comparing two cases within the same sentencing ranges, the magnitude of the effect of changing sentencing ranges on these two cases depends on the current position of the case within the sentencing range. This follows from the same intuition as the first implication. The closer to the upper limit of the sentencing range, the higher the effect of an increase in the upper limit. This effect can be attenuated or amplified by sentencers' practice not to exploit the full sentencing range. For example, Italian judges counter the punitive tendencies of the legislator by imposing sentences close to the lower limit of the sentencing range (Corda, 2016), limiting the severity mechanism caused by increasing the upper limit of the sentencing range.

Third, the reference effect is likely higher if new cases are similar in terms of relative

seriousness. Imagine a legislative change increasing the 50k threshold to 100k for thefts, while retaining the 500k threshold. Cases just above 100k will be influenced more strongly than those close to the 500k threshold. Similarly, cases just below 50k will be sentenced more differently than those with minimal damage. We suspect that if less serious cases are added into a subsection, then the initially least serious cases will become relatively more serious by more than the initially most serious cases and *vice versa*.

Studying the average effect masks substantial heterogeneity in prosecutors' behavior. Figure 2.7 and Figure 2.8 in the Appendix show histograms of all sentences recommended by individual prosecutors for each scenario. The distribution of recommended sentences within each scenario exhibits a large variation. Strikingly, in the majority of scenarios, they chose sentences close to both the lower and the upper limits of sentencing ranges. Furthermore, the figures show that prosecutors tend to round the length to years, which is in line with previous research by Dhami et al. (2020); Pease and Sampson (1977). Quantity thresholds are thus not the only source of sizable sentencing disparities in the Czech criminal justice system.

An experimental design seems the most appropriate for future research into nuanced roles of thresholds' effects. Defendants, police and prosecutors respond to quantity thresholds (Bjerk, 2005, 2017b; Lepage, 2020; Travova, 2019), rendering any real-data sentencing study complicated. The studied cases need to be such that it is difficult for defendants and state representatives to tailor their response to quantity thresholds, which can be best achieved in an experimental setting.

We further propose a novel measure of justice. We formalize the notion of justice such that similar cases should be treated similarly and different cases differently and introduce a formal definition of a *just* sentence. In particular, a sentence is considered *just* if it is harsher than a sentence imposed for a more severe case, but not unreasonably harsher. The proposed empirical strategy of quantification has two applications. First, the measure of the *Type I injustice* allows us to quantify (perceived) differences between two criminal cases. If the two cases are perceived by sentencers as identically serious, then the measure of the *Type I injustice* converges to $\frac{1}{2}$. The more they differ, the smaller the *Type I injustice* is. Importantly, since *Type I injustice* does not depend on the tolerance parameter and has no units, it is easily applicable and comparable across different problems. Second, the measure of the *Type II injustice* provides policy makers with a simple tool – easily simplified into a number – to compare and evaluate sentencing disparities caused by excessive harshness in sentencing.

We believe that both our research question and our results contribute to a rather general question of how to structure criminal codes and offense subsections. We do not see our results as supporting an argument that thresholds are generally wrong and sanctions need to be (linearly) proportional to measurable harm (drug possessed or damage caused). Rather, we see our results as evidence that can help to guide a discussion about the structure of criminal codes. Simply, the discussion is not whether thresholds are useful or harmful, but rather how they should be implemented and how the subsections and ranges should be constructed. In particular, what classifying variables should divide offense subsections, into how many subsections, and should some classifying variables be quantifiable? To what extent should sentencing ranges for individual subsections overlap? To provide answers to these questions, it seems necessary to understand how offense subsections and thresholds influence sentencing decisions.

2.6 Appendix A

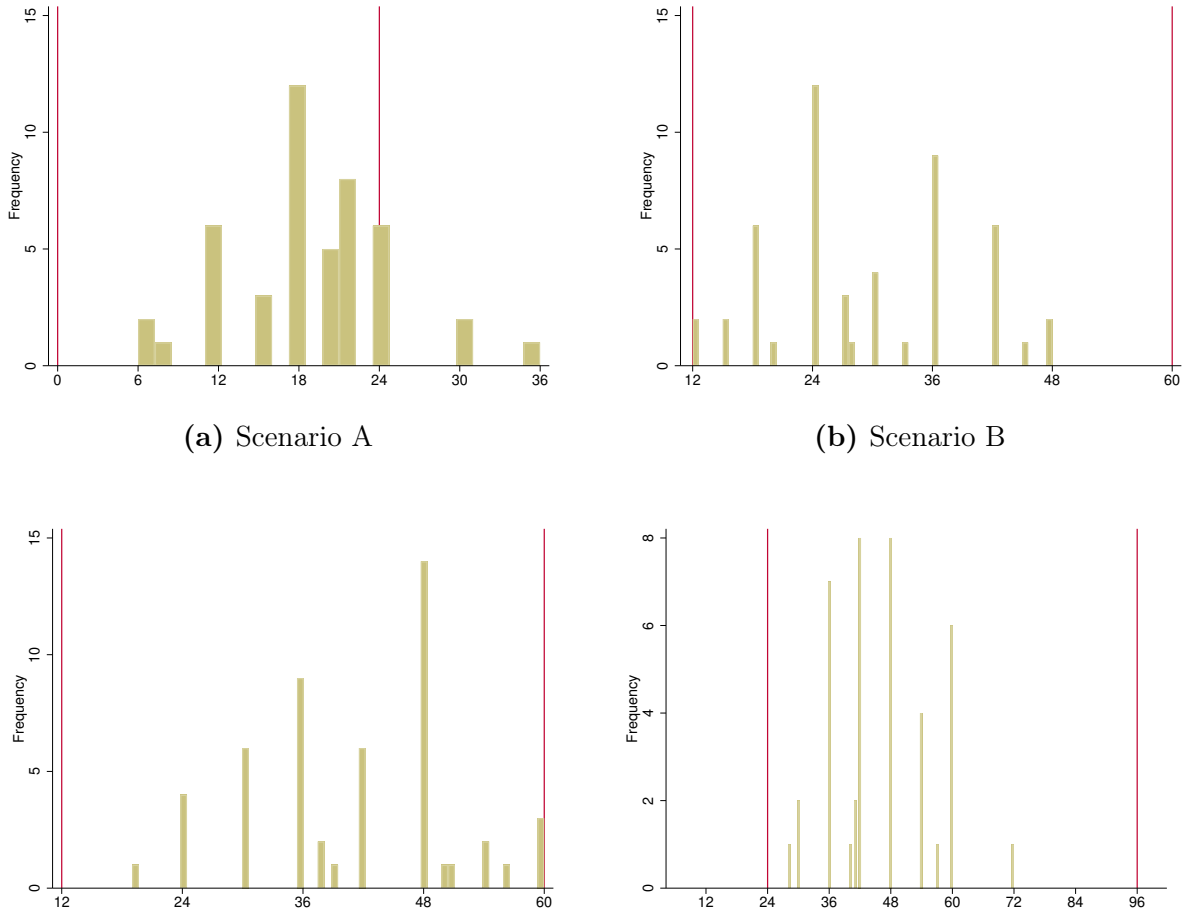
We show that equation T 1.1 holds. To simplify notation, label two cases A and B and define three differences $\Delta G = G(B) - G(A)$, $\Delta\rho^- = \rho^-(B) - \rho^-(A)$, and $\Delta\rho^+ = \rho^+(B) - \rho^+(A)$. Then, according to Definition 1, the difference in sentences $\Delta s = s(B) - s(A)$ equals to

$$
\begin{aligned}
s(B) - s(A) &= \rho^-(B) + G(B) * \big(\rho^+(B) - \rho^-(B)\big) - \big(\rho^-(A) + G(A) * \big(\rho^+(A) - \rho^-(A)\big)\big) \\
&= \rho^-(B) + G(B) * \big(\rho^+(B) - \rho^-(B)\big) - \\
&- \big((\rho^-(B) - \Delta\rho^-) + (G(B) - \Delta G) * \big((\rho^+(B) - \Delta\rho^+) - (\rho^-(B) - \Delta\rho^-)\big)\big) \\
&= \Delta\rho^- * (1 - G(B) + \Delta G) + \Delta\rho^+ * (G(B) - \Delta G) + \Delta G * (\rho^+(B) - \rho^-(B)) \\
&= \Delta\rho^- * (1 - G(A)) + \Delta\rho^+ * (G(A)) + \Delta G * (\rho^+(B) - \rho^-(B))
\end{aligned}
$$

□

2.7 Appendix B

Figure 2.7: Theft: Individual Sentences Recommended by Prosecutors



(a) Scenario A

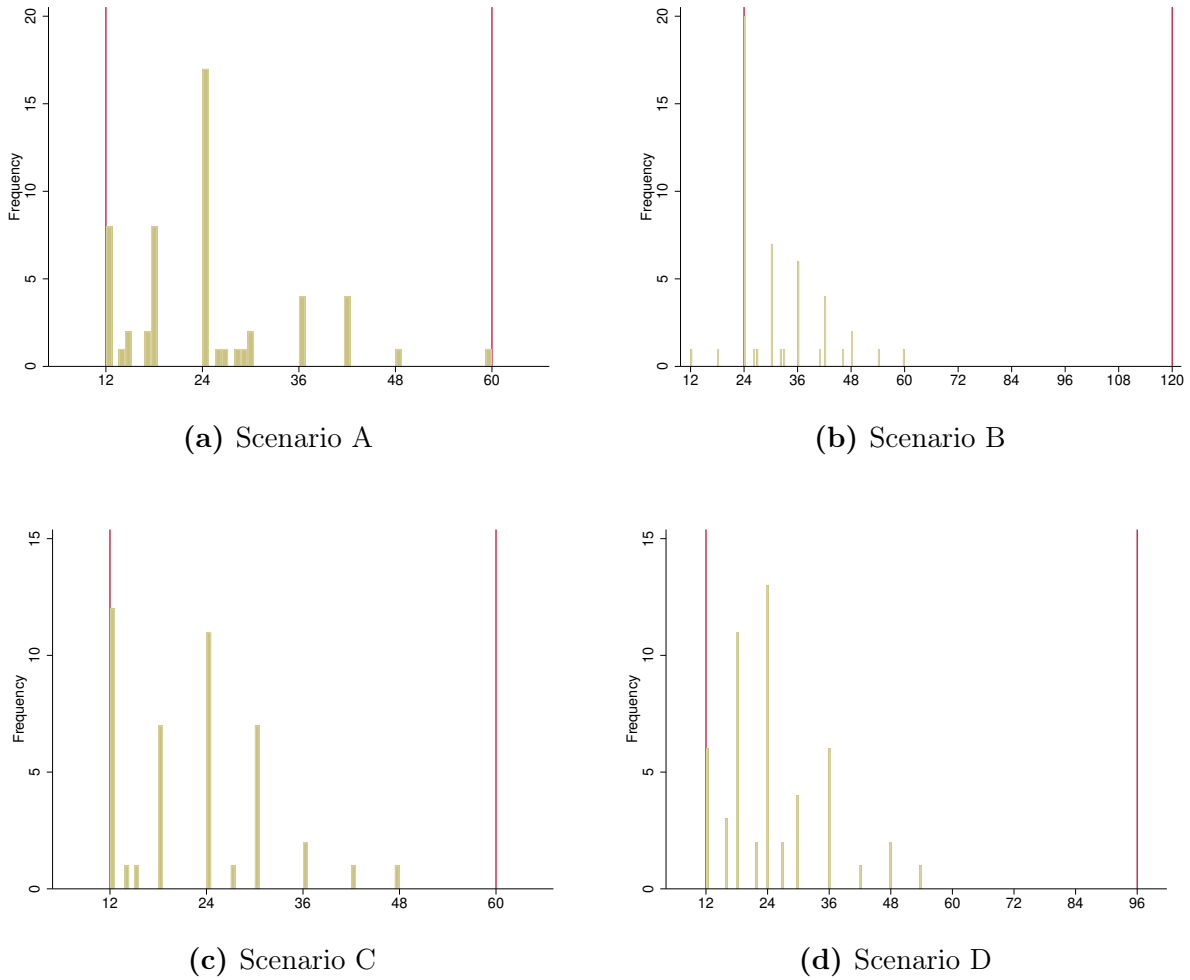
(b) Scenario B

(c) Scenario C

(d) Scenario D

Notes: Each panel represents a histogram of individual recommended sentences in the case of theft. In scenario A, prosecutors recommended sentences in a case with damage of CZK 48,283. In scenario B, the prosecutors recommended sentences in a case with damage of CZK 51,283. In scenario C, prosecutors recommended sentences in a case with damage of CZK 487,092. Finally, in scenario D, prosecutors recommended sentences in a case with damage of CZK 508,213. Red vertical lines mark the upper and lower limits of the corresponding sentencing range. See Table 2.1.

Figure 2.8: Drug Possession: Individual Sentences Recommended by Prosecutors



(a) Scenario A

(b) Scenario B

(c) Scenario C

(d) Scenario D

Notes: Each panel represents a histogram of individual recommended sentences in the drug possession case. In scenario A, prosecutors recommended sentences in a case with 147.8 g. In scenario B, the prosecutors recommended sentences in a case with 151.8g. In scenario C, prosecutors recommended sentences in a case with 147.8 g and a composition of more serious cases. Finally, in scenario D, prosecutors recommended sentences in a case with 147.8 g and higher upper limit of the sentencing range. Red vertical lines mark the upper and lower limits of the corresponding sentencing range. See Table 2.3.

Table 2.6: Descriptive Statistics

Panel A: Drug Possession

	A	B	C	D	balance test	not in exper.	H_0
Number of Observations	58	50	44	54		1049	
Male	41 %	42 %	34 %	52 %	0.361	55 %	0.001
Communist Party	10 %	18 %	10 %	15 %	0.537	14 %	0.841
Age	46.3	47.8	45.2	47.3	0.660	49.1	0.002
Tenure Exam	18.5	19	16.6	20.3	0.426	20.4	0.066
Tenure Oath	16.2	16.6	13.8	16.4	0.334	17.1	0.071
Alma Mater							
Brno	38 %	38 %	32 %	38 %	0.804	34 %	0.529
Prague	34 %	40 %	43 %	46 %	0.574	42 %	0.662
Plzen	16 %	8 %	11 %	2 %	0.087	9 %	0.799
Olomouc	5 %	4 %	7 %	4 %	0.895	7 %	0.840
NSZ	10 %	6 %	9 %	9 %	0.880	4 %	0.012
VSZ	3 %	2 %	14 %	2 %	0.023	8 %	0.147
KSZ	16 %	14 %	20 %	17 %	0.860	24 %	0.025
OSZ	70 %	78 %	57 %	72 %	0.153	64 %	0.114

Panel B: Theft

	A	B	C	D	balance test
Number of Observations	46	51	51	46	
Male	43 %	27 %	51 %	54 %	0.067
Communist Party	20 %	16 %	6 %	13 %	0.353
Age	48	46.6	47	46.2	0.836
Tenure Exam	19.8	19	18.4	18.5	0.960
Tenure Oath	16.7	18.9	15.9	15.2	0.860
Alma Mater					
Brno	37 %	37 %	37 %	35 %	0.880
Prague	39 %	41 %	37 %	48 %	0.497
Plzen	7 %	14 %	10 %	4 %	0.401
Olomouc	9 %	0 %	8 %	4 %	0.217
NSZ	11 %	14 %	4 %	9 %	0.308
VSZ	4 %	2 %	4 %	9 %	0.505
KSZ	15 %	14 %	20 %	20 %	0.819
OSZ	70 %	71 %	73 %	63 %	0.882

Notes: The H_0 column reports the p-value of the two-sided t-test under the null that the sample value of participants equals those who do not participate.

Balance test represents p-value of F-test under the null that there is no difference between the groups.

Introduction Screen[16] Welcome!

We ask you to participate in a scientific study, in which we explore sentencing decisions.

On the following screens we present two hypothetical criminal cases. Presented legal provisions, according to which you will decide, might not correspond to the current legal provisions. Your task will be to recommend the length of the sentence.

After that we will present you a questionnaire. Your answers are and will remain fully anonymous and will be used only for research purposes. Your participation should not take up more than 10 minutes of your time.

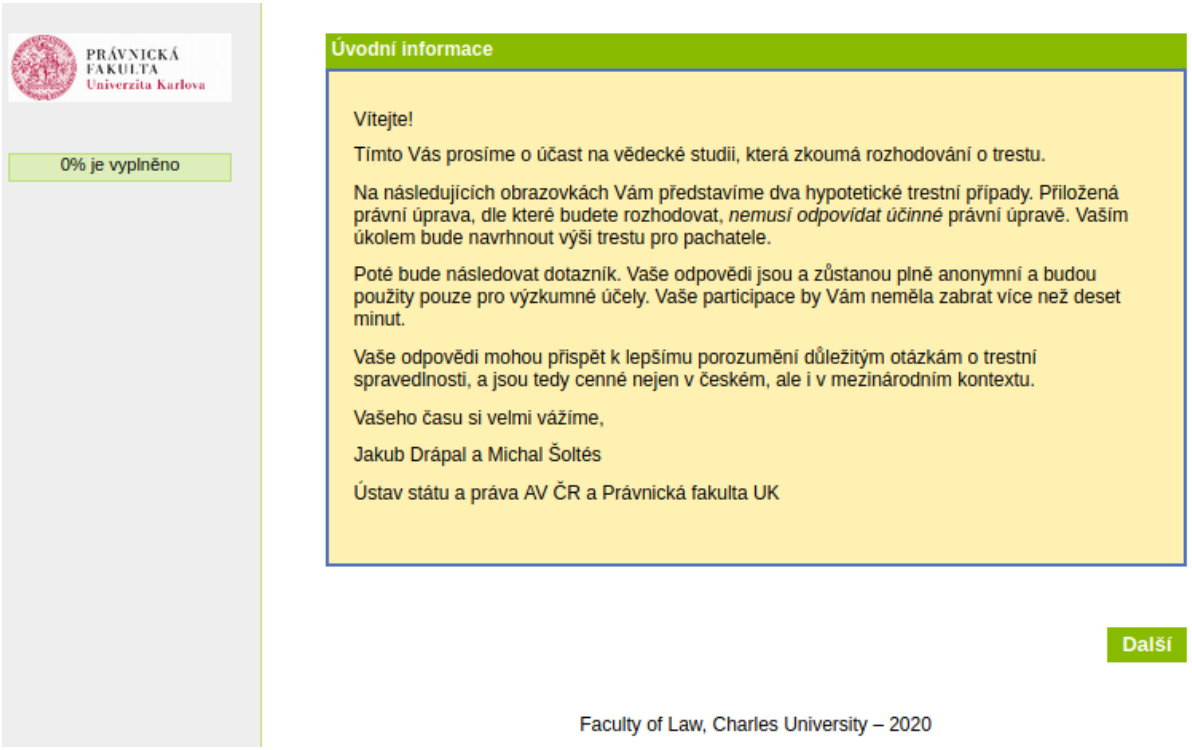
Your answers might contribute to the better understanding of important criminal justice issues that are significant not only for the Czech Republic, but for the international audience as well.

We highly value your time.

Jakub Drápal and Michal Šoltés
Institute of State and Law, the Czech Academy of Sciences and Faculty of Law, Charles University

[16] See figure 2.9

Figure 2.9: Vignettes: Introduction Screen



Drug distribution[17] Pavel Nový (born 14. 5. 1984, unemployed, resident of Chomutov) was arrested by policemen in front of a dance club in Chomutov while selling methamphetamine, which he bought a day earlier in Prague. According to a lab report, the amount of methamphetamine found on the offender (in his pockets and in the car parked in front of the dance club) contained [amount] of pure substance of methamphetamine.

A month prior to the arrest Mr. Nový lost his job, with financial troubles ensuing. While looking for job in Prague, he came across an old acquaintance who offered him a one-time possibility to earn some money by selling drugs.

In the last 10 years, Nový was thrice sentenced for distribution of marijuana. Probation period of the last sentence elapsed four years ago, during which he was not found in breach of the conditions. Two and a half years ago he was sentenced for burgling several residential houses and apartments to probation; he was not found in breach. Half a year ago, he was sentenced for a small theft in a supermarket to a community service sentence, which he carried out.

Even though he did not cooperate with the police in the beginning, he pleaded guilty and there is no doubt regarding his guilt as well as the legal classification of the offense.

The head of your prosecution office told you during a preliminary discussion that you should recommend a non-suspended prison sentence based on the offender's criminal record and the amount of drugs found on him. He left the decision on the length of the non-suspended prison sentence entirely up to you.

Offense Section: Unauthorised Production and other Disposal with Narcotic and Psychotropic Substances and Poisons

(1) Whoever produces, imports, exports, transports, offers, provides or sells or otherwise arranges for another or handles for another narcotic or psychotropic substances, products containing narcotic or psychotropic substances, precursors

[17] **See figure 2.10**

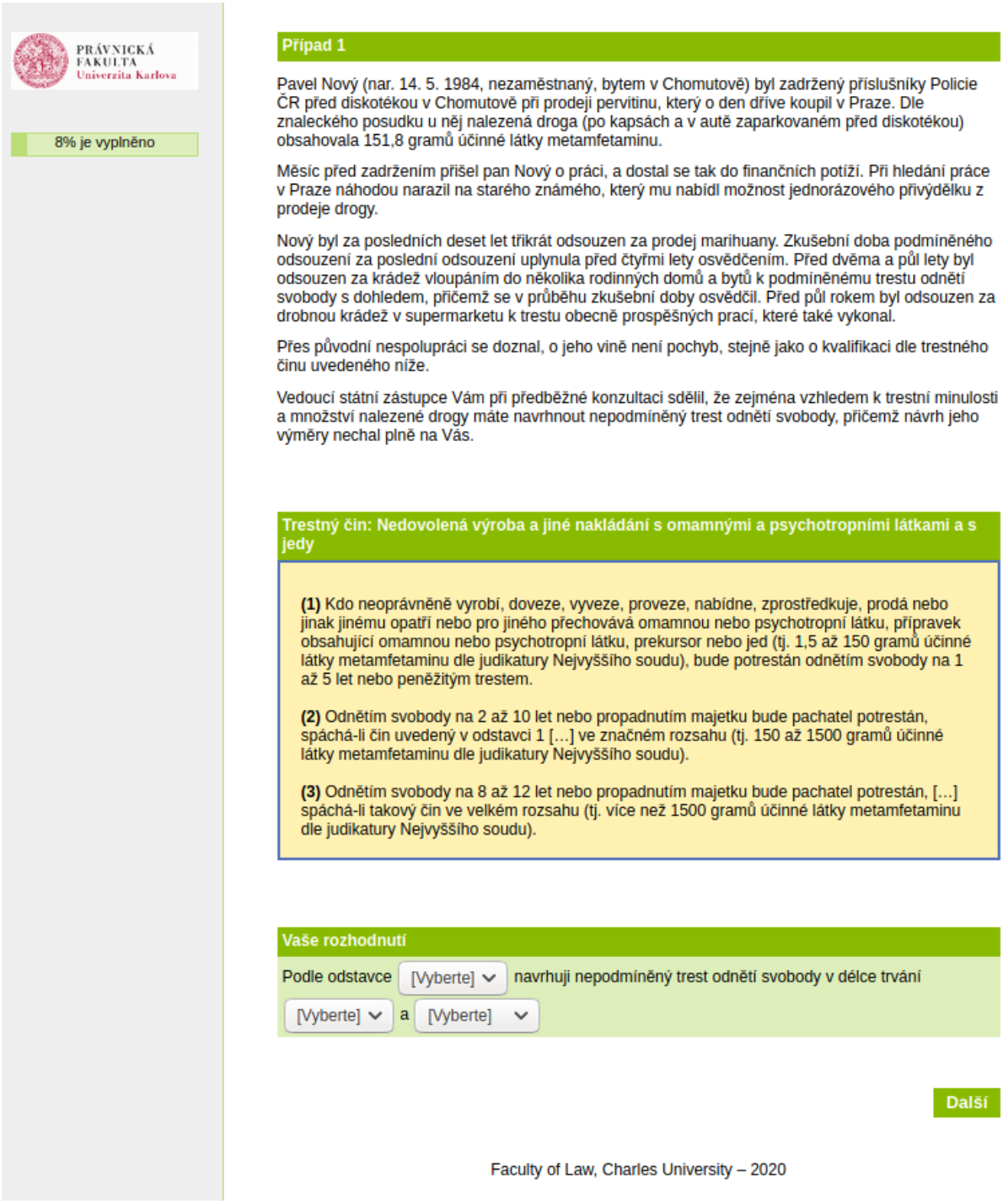
or poisons (meaning 1.5-150 grams of pure substance of methamphetamine according to the jurisprudence of Czech Supreme Court), without an authorisation, shall be sentenced to imprisonment for one to five years or to a pecuniary penalty.

(2) An offender shall be sentenced to imprisonment for two to ten years or to confiscation of property, if he/she commits the act referred to in Sub-section (1) [...] in a considerable extent (meaning 150-1500 grams of pure substance of methamphetamine according to the jurisprudence of Czech Supreme Court).

(3) An offender shall be sentenced to imprisonment for eight to twelve years or to confiscation of property, if he/she [...] commits such an act in a large extent (meaning more than 1500 grams of pure substance of methamphetamine according to the jurisprudence of Czech Supreme Court).

Your decision According to the subsection [Choose] I recommend a non-suspended prison sentence in the length of [Choose] years and [Choose] months.

Figure 2.10: Vignette 1: Drug Distribution



Vignette 2: Theft[18] Karel Pokorný (born 5. 10. 1978, unemployed), visited his parents in a morning before they left for work to help them update software on their home computer. He stayed in their apartment even after they left it for their job to finalize the update.

He took advantage of the fact that his parents were logged into their internet banking and that a text message with confirmation code is sent to their family cellphone, which they had left at home. He entered a payment order via which he transferred all of his parents' money to his own account and he confirmed the payment by a confirmation code. He thus caused damage to his parents of [amount]. He gambled away all of the money in a local casino on slot machines.

Pokorný has 8 previous convictions for fraud, embezzlement and theft, of which he was sentenced thrice for fraud in the last five years. He served the last sentence (non-suspended prison sentence of two years) two months prior to this event.

He plead guilty, he did not compensate his parents and there is no doubt regarding his guilt as well as the legal classification of the offense.

The head of your prosecution office told you during a preliminary discussion that you should recommend a non-suspended prison sentence based on the offender's criminal record. He left the decision on the length of the non-suspended prison sentence entirely up to you.

Offense Section: Theft

(1) Whoever misappropriates a thing of another by taking possession of it, and thus causes damage not insignificant on property (meaning CZK 5,000-50,000) of another shall be sentenced to imprisonment for up to two years, to prohibition of activity or to confiscation of a thing or other asset value.

[18] **See figure2.11**

...

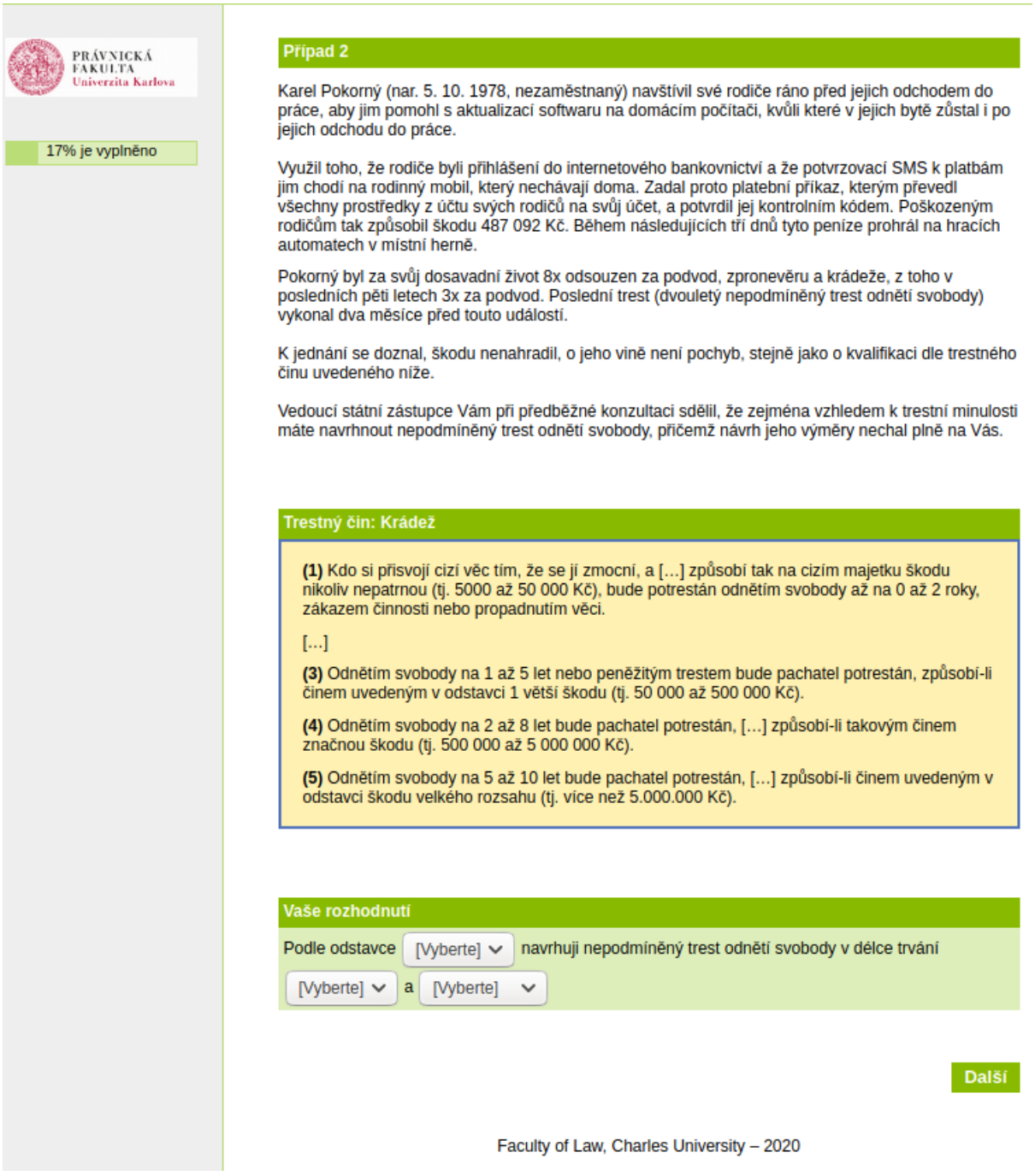
(3) An offender shall be sentenced to imprisonment for one year to five years or to a pecuniary penalty, if he/she causes larger damage (meaning CZK 50,000-500,000) by the act referred to in Sub-section (1).

(4) An offender shall be sentenced to imprisonment for two to eight years, if he/she [...] causes substantial damage (meaning CZK 500,000-5,000,000) by such an act.

(5) An offender shall be sentenced to imprisonment for five to ten years, if he/she [...] causes by the act referred to in Sub-section (1) extensive damage (meaning more than CZK 5,000,000).

Your decision According to the subsection [Choose] I recommend a non-suspended prison sentence in the length of [Choose] years and [Choose] months.

Figure 2.11: Vignette 2: Theft



Chapter 3

Consequences of Inconvenient Information: Evidence from Sentencing Disparity

3.1 Introduction

Public sector institutions have been repeatedly found reluctant to disclose inconvenient information, even if not doing so may jeopardize public health and safety and undermine principles of modern democracies. Censorship of information after the Chernobyl catastrophe by Soviet propaganda and repeated efforts to cover-up the spread of diseases in China are two prominent examples. In these examples, and in many other cases, the aversion to disclose information to the public was supposedly motivated by concerns that the information might reveal the incompetence and systematic failure of responsible authorities. And this, in turn, might lead to public distrust in the system and its institutions.[1]

In many cases, the concerns may be valid. In general, information about the performance and competence of public institutions to deliver on their responsibilities (e.g., public health ensured by health officers) affects public trust in institutions. Importantly, the information is likely to shape citizens' behavior with economic and social consequences. For example, Acemoglu et al. (2020) document that providing the general public with pos-

[1]For example, the Associated Press (2020) reports on the COVID-19 outbreak in China as follows: *"In Wuhan, local leaders were accused of telling doctors in December not to publicize the spreading virus in order to avoid casting a shadow over the annual meeting of a local legislative body. As the virus spread, doctors were ordered to delete posts on social media that appealed for donations of medical supplies. That prompted complaints authorities were more worried about image than public safety."*.

itive information about state courts' performance (reduced delays) in Pakistan changed citizens' attitude and increased the likelihood of using state courts instead of relying on informal institutions for dispute resolution. Since the asymmetric impact of negative and positive information has been documented in many domains (Eil and Rao, 2011; Coutts, 2019; Galil and Soffer, 2011; Moutsiana et al., 2013), one may wonder whether disclosing negative – and for the public institution "inconvenient" – information would lead to the opposite: public distrust and avoidance of such institutions.

To provide empirical evidence on the consequences of inconvenient information regarding the performance of public institutions, I conduct a survey experiment studying how citizens respond to information about sentencing disparities among judges in the Czech Republic. The fact that judges' characteristics (e.g. propensity to incarcerate) affect sentencing decision is arguably one of the most worrying signals regarding the performance of judicial systems, which challenges the formal rules of equality before the law, and the clear, stable, and predictable application of law.[2] In the experiment, 2,410 participants were randomized into a treatment and a control group. The groups were provided with varying yet not deceptive information about sentencing disparity among judges at regional courts in the Czech Republic. The treatment group was informed about sentencing decisions at a court where judges differ in their sentencing practice (sizeable sentencing disparity), while the active control group was informed about a court where all presented judges tend to decide consistently (negligible sentencing disparity).

The core of the information treatment consists of shares of cases in which convicted offenders were sentenced to community service instead of other types of punishments (e.g., incarceration) for one of the most frequent crimes in the Czech Republic - *failure to pay alimony*. In a between-subjects design, I then measure the effect of the information treatment on: (i) declared trust in several institutions, including the judicial system; (ii) court-related behavior, such as the willingness to apply to courts and the demand for alternative dispute resolution; and (iii) policy preferences regarding the judicial system, including a willingness to become actively engaged in addressing the sentencing disparity

[2]Sentencing disparity is not unique to the Czech judicial system. In fact, it has been documented in many other judicial systems worldwide and extensively discussed by scholars. Mainly in criminology, but also in other related fields, a lack of consistency in sentencing is an established fact. Disparities have been documented along different dimensions: (i) within judges across time; (ii) between judges in a single jurisdiction; and (iii) between jurisdictions (Sporer and Goodman-Delahunty, 2009b). Many scholars have even leveraged the different practice (leniency) of judges as a source of quasi-exogenous variation to provide causal estimates of incarceration on various outcomes (see, for example, Kling, 2006; Di Tella and Schargrodsky, 2013; Dahl et al., 2014; Aizer and Doyle, 2015).

by signing a petition.

Disclosure of inconvenient information about sentencing disparity did not have adverse consequences of citizens losing faith in the formal institutions. In particular, the treatment did not lead to public distrust in the judicial system or any other intuitions (e.g. the police). Furthermore, I find no effect on the intention to avoid the formal judicial system and on demand for alternative dispute resolution. Instead, it seems that respondents in the treatment group are more likely to become involved in the solution to the problem. The treatment motivated respondents to sign a petition that calls on politicians to suggest specifying sentencing principles that would assist judges in their sentencing decisions and thus limit the sentencing inconsistency. By the most conservative estimates, the treatment increases the share of respondents willing to sign the petition by 3.3 percentage points (5.8%). Additionally, subjects exposed to the treatment information found fairness of the judicial system a more important policy goal than the other subjects.[3]

To understand who is likely to react to the inconvenience information, I made use of the fact that mothers (in the data identified as female respondents with at least one child regardless of the age of that child) are arguably more sensitive to the incompetence of the judicial system in the case of *failure to pay alimony*. If after a divorce, a father refuses or is unable to pay alimony for his children, a mother is left to cover necessary expenditures alone. Unfortunately, this happens frequently.[4] As a result, divorced mothers are often left in a complicated financial situation with little help from the government and any other institutions. Even if NGOs provide help and assistance, they focus on legal consultation regarding suing the defaulter rather than providing financial support. Consequently, mothers are likely to be more sensitive to the treatment information than any other groups, as mothers are the most vulnerable.

[3]The results qualitatively correspond to the reaction of the general public to information and a video of the death of George Floyd. The general public has became undoubtedly more interested in the issue of racism documented by the Black Lives Matter protests and, for example, by online search (Barrie, 2020). A survey conducted 2 months after George Floyd's death suggests that most Americans support major (58%) and minor (36%) changes in policing; however, only 15% support the idea of abolishing police departments (Crabtree, 2020). Similarly, Vaughn et al. (2021) also find that the public is significantly more supportive of reforming the police than defunding or abolishing. Additionally, Philonise Floyd, George Floyd's brother, called on lawmakers to make law enforcement part of the solution, not the problem during a House Judiciary Committee hearing to discuss police brutality and racial profiling, in Washington, on June 10, 2020.

[4]There is no exact figure of how often this happens, but the fact that the *failure to pay alimony* is one of the three most frequent crimes, and only a fraction of the cases go to court, suggests that this is a sizeable problem.

The results suggest that the treatment effect indeed varies by respondents' characteristics. The overall willingness to act on the information and sign the petition is driven exclusively by mothers. Once I allow the treatment effect to vary by the mother-status of respondents, mothers in treatment groups are 10.7 percentage points more likely to sign the petition than mothers in the control group and 7.7 percentage points more likely than non-mother treated respondents. Under less conservative specifications, the effect is even larger. A similar pattern is discernible in the reported trust in the judicial system. Importantly, the potential negative consequences of the effect on reliance on the judicial systems are not visible even among mothers. Overall, zooming in on the most sensitive group of respondents makes the results only stronger: Information about sentencing disparity did not lead to avoidance of the judicial system. Instead, it motivates the respondents to become engaged and demand improvement. I found no evidence that the treatment effect would depend on respondents' prior beliefs whether the judicial system works well or not.

Previous academic literature has studied the consequences of publishing information under different conditions. First, scholars have studied the effects of information about the performance of private firms (see, for instance, Beyer et al., 2010). This project differs from that stream of literature, as a reaction to the disclosure of firms' performance usually materializes through market mechanisms and affects the firms' valuation, which is virtually impossible in the case of public institutions. Second, previous discussions in economics regarding information disclosure by public institutions focused on precision of the information and, in particular, on a trade-off between timely but noisy information and slow but more accurate information regarding volatile economic statistics such as GDP (Morris and Shin, 2002) and on communication strategies of central banks as a monetary policy tool (Blinder et al., 2008). This literature thus differs in the nature of the information and its goals. Information about the state of the economy and intended monetary policies aim to increase transparency and form market expectations. As long as the communication of central banks does not lead to questioning the competence of the central bank to deliver on monetary policy goals, this project is less related.

This project is more relevant to the stream of literature devoted to consequences of publishing the performance indicators of hospitals (Smith et al., 2009; Ketelaar et al., 2011), as public health is often (co)financed through public budgets with regulated prices that limits the scope for market mechanisms. However, since patients are generally allowed to choose which hospital to use, even regulated market mechanisms work and

patients prefer better performing hospitals. As a result, the consensus in the literature suggests that publishing information has led to an improvement in under-performing hospitals (Hibbard et al., 2005). This project clearly differs, as offenders cannot generally choose which court to attend and courts are not financed according to the number of cases decided; market mechanisms do not apply at all.

This project shares several features with a recent paper by Acemoglu et al. (2020). The authors show that information about reduced delays in state courts increases the reported likelihood of using formal courts instead of non-state institutions (*Panchayats*) in rural Pakistan. Their results suggest higher information sensitivity compared to mine. Importantly, the studies differ diametrically in their context: (i) Information provided by Acemoglu et al. (2020) is viewed as positive, whereas mine as negative. Since the previous literature documented an asymmetric reaction to negative and positive news in many domains of human behavior (Eil and Rao, 2011; Coutts, 2019; Galil and Soffer, 2011; Moutsiana et al., 2013), it is likely an important difference in this setting too. (ii) Their project was conducted in a rural area in Pakistan where households access the court system frequently[5] and thus are aware of, and are arguably more sensitive to, the performance of the judicial system. In my setting–a standard European democracy–awareness about courts' performance is less widespread, and respondents are less experienced in the judicial system and courts' practice. Only 10% of respondents in this study reported that they had had sizeable experience with the judicial system (first-hand and/or through people they know well, e.g., family).[6] (iii) The judicial systems in Pakistan and the Czech Republic enjoy different levels of public trust. According to Eurobarometer (2018), 43% of respondents in the Czech Republic tend to trust the judicial system. That is 8 percentage points fewer than the average of the EU28, yet still comparable with most developed countries. Conversely, in Pakistan, the state institutions suffer from a lack of trust (Jackson et al., 2014; Cheema et al., 2017).[7] All three aspects likely contribute to different information sensitivity.

Finally, I contribute to a broad literature studying institutional trust. Unlike inter-

[5] "In our survey one in every five households report that they have accessed the court system in the last three months" (Acemoglu et al., 2020, p.7).

[6] A lack of knowledge about how the judicial system works among the general public seems to be common in the European democracies. For example, according to Chapman et al. (2002), the British Crime Survey (BCS) and other surveys have shown that the public is poorly informed about crime and the operation of the criminal justice system.

[7] "Pakistan is an ideal setting for such an investigation because of the well-recognized weakness of state institutions and the associated low levels of access to and trust in the state." (Acemoglu et al., 2020, p.1).

personal trust, which is usually elicited using the trust game, institutional trust is more challenging to measure. Regrettably, there is no evidence that the properties of interpersonal trust are automatically transferable to different domains. Carlsson et al. (2018) and Alesina and La Ferrara (2002) find a low correlation between generalized trust and trust in various institutions (the government, the police, the judicial system). Many scholars studying institutional trust rely on data about declared institutional trust and provide correlational evidence (e.g., Grönlund and Setälä, 2012). Evidence on the causal effect of interventions on institutional trust is generally rare. Two studies estimate the causal effect of the perceived quality of public institutions on institutional trust by analyzing procedural justice protocol and trust in the police. Murphy et al. (2014) finds that when police officers followed an experimental protocol — that focused on voice, neutrality, trustworthiness, and respect — during a control, then drivers in Australia reported higher trust in the police. However, using a similar experimental design in Scotland, MacQueen and Bradford (2015) failed to replicate the effect of an increase in trust. A similar question of whether judicial system transparency affects institutional trust was studied by Grimmelikhuijsen and Klijn (2015). In their field experiment, respondents were invited to watch a TV series about a district court in the Netherlands that allowed the public to watch judges' daily work on real cases. The authors report that watching the TV series increased the declared level of trust in judges. The treatment, however, conveys different information. Information on the day-to-day practice on several cases can barely reveal (in)consistency in sentencing among judges. I extend this stream of literature by estimating causal effects of information about public institutions' performance on declared institutional trust and other measures of intended behaviors related to the trust.

The rest of the chapter is organized as follows. The following section introduces the design of the experiment, and the outcomes studied. Next, I discuss the results with attention to the heterogeneous treatment effect by the mother-status. Finally, before I conclude, I remark on the interpretation and implications of the results.

3.2 Design of Survey Experiment

To conduct the survey experiment, I partnered with Behavio, a private company administrating a panel of regular respondents. Respondents were invited by email to take part in an online survey about courts and justice. 2,410 respondents completed the survey. In

addition to the data collected in the experiment, I have basic demographic characteristics of respondents collected in previous surveys. Except for the final task, the experiment was run on a platform of Behavio familiar to the respondents.

The experiment consisted of four stages. Upon starting the survey, respondents were asked three questions regarding their prior attitude to the judicial system and their previous experience. Next, respondents were randomized into the treatment and the active control groups and were presented with the corresponding information. After the treatment phase, respondents were asked to complete five tasks and questions. The final task consisted of reading and signing a petition posted on a different website. Respondents interested in the petition had to leave the Behavio platform. The order of tasks and questions in the first (prior attitude) and the third stages (five tasks) were randomized at the individual level. The full script is available in the appendix.

3.2.1 Attitude

The first stage aimed to understand the respondents' initial attitude towards the judicial system. Respondents were asked to what extent on a four-level scale they agree with two statements: (i) "Depending on the judge, similar cases can be sentenced differently"; (ii) "Overall, the judicial system in the Czech Republic works well." Additionally, I asked how experienced with the judicial system they and/or people close to them are.

3.2.2 Treatment

I provided respondents with varying yet not deceptive information about sentencing disparity among judges within a regional court. The treatment group was informed about a court with high sentencing disparity among judges, while the active control group about a court with a negligible sentencing disparity among judges. The source of the variation comes from sentencing disparities among judges at different courts. At some regional courts, judges significantly vary in their sentencing patterns, while at others, judges exhibit indistinguishable sentencing patterns. The information relies on variation within a given court, rather than between courts, as some of the regional disparities in sentencing are justifiable and do not represent the intended variation.[8]

[8]For example, a driving disqualification in a city with functional public transportation is arguably a more lenient punishment in terms of economic and social consequences than the same type of punishment in regions at the foothills of mountains with limited public transportation. These and similar considerations may lead to some desired sentencing disparity across regions.

Data about sentencing decisions are complicated and multidimensional,[9] which makes it complicated to convey an understandable message. I rely on one of the most common offenses in the Czech Republic – *failure to pay alimony*[10] – and present shares of cases in which a judge sentences a convicted person to community service as the primary punishment. The offense of *failure to pay alimony* satisfies four criteria needed for the treatment to be based on credible information: (i) there are enough observations so that I can provide aggregate statistics based on at least 80 cases per judge (over the three year period of 2016-2018); (ii) compared to other offenses, in the objective elements of a crime, *failure to pay alimony* is a homogenous crime; (iii) while it is not part of the information provided in the treatment, the differences highlighted in the treatment are statistically significant; (iv) since it is a general type of crime, cases are assigned to judges at random.

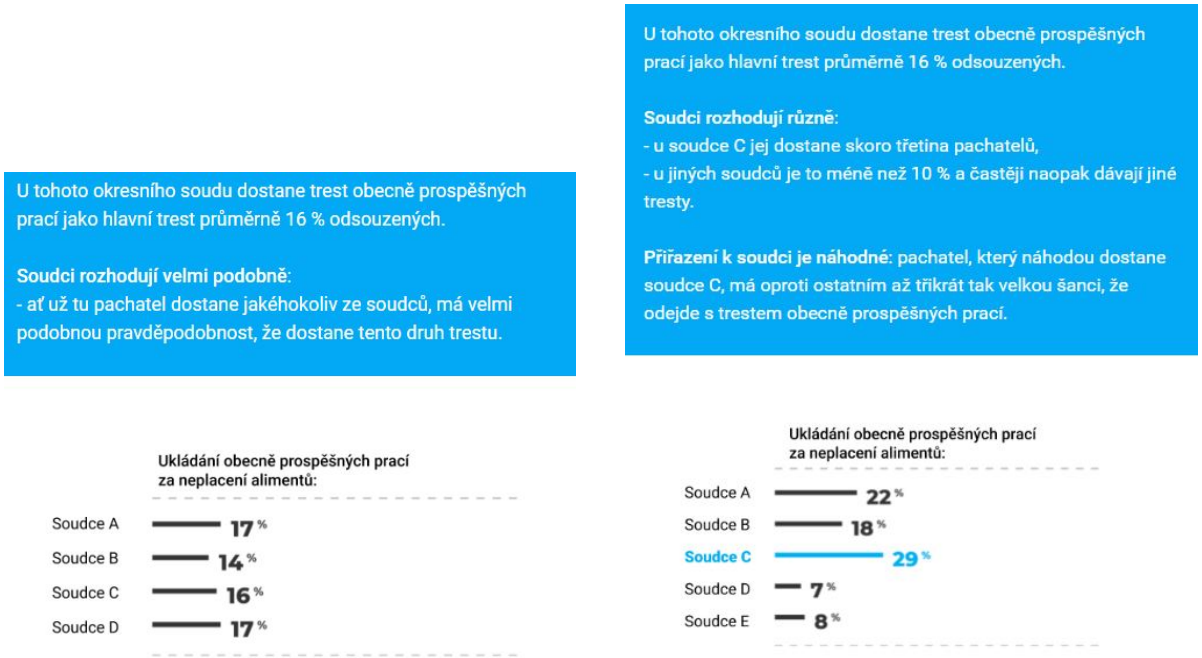
The treatment and the control slides present the corresponding information in a way that the news would do. The slide's core is a simple, self-explanatory bar graph accompanied by a few additional pieces of information providing an interpretation of the graph. In particular, the treatment slide consists of a bar graph showing shares of cases in which the convicted criminals were sentenced to community service by different judges (22%, 18%, 29%, 7%, and 8%), the headline says: "Judges sentence differently." Next, the slide explains that judge C (29%) sentenced almost a third of the convicted offenders to community service, whereas, for some, it is less than 10% and instead, they impose different types of punishments. Finally, the slide highlights that cases are assigned at random and that being assigned to judge C implies up to a threefold higher probability of being sentenced to community service. In the control group, the slide shows a bar graph with shares of cases that were sentenced to community service by different judges (17%, 14%, 16%, 17%). The headline says: "Judges sentenced very similarly." The control slide further explains that regardless of the judge assigned, a convicted offender has a very similar probability of being sentenced to community service. Respondents in both groups are informed that the figures are based on actual sentencing decisions of judges at one of the Czech regional courts, but they do not know which one. Figure 3.1 shows the screen shots of the control and treatment slides.[11]

[9] One has to consider different offenses and their subsections, different types and extent of punishments, and combination of more types of punishments.

[10] Formally, the crime is called Section 196 Negligence of Mandatory Support.

[11] Once respondents finished the experiment, Behavio sent them a debriefing letter that explains that the information presented represents only one regional court and the situation may differ in different courts. The debriefing letter can be found in the online appendix.

Figure 3.1: Control and Treatment Slides



(a) Panel A: Control Slide

(b) Panel B: Treatment Slide

Notes: The control slide consists of a graph showing the frequency of community service used by four different judges. The treatment slide consists of a graph showing the frequency of community service used by five different judges. Both graphs is accompanied by a brief explanation of the graph.

3.2.3 Experimental Outcomes

The collected outcomes are classified into three main categories: (i) declared institutional trust; (ii) reliance on the judicial system; and (iii) policy preferences.

Institutional Trust To measure institutional trust, I adopt standard survey questions of declared institutional trust similar to those used by international institutions such as the World Values Survey (WVS) and the Eurostat. In particular, respondents were asked to indicate their trust level on a scale of: *a great deal*; *quite a lot*; *not very much*, and *none at all* towards four different institutions. One of the institutions was the judicial system. The choice of the others was related to the judicial system. The closest institution to the judicial system is the police, as police officers often cooperate on criminal cases. The next institution is the government, which is responsible for a functioning judicial system[12]; and finally, the public broadcasting service, which can be viewed responsible for the lack of information about the sentencing disparity.

Reliance on the Judicial System Next, I propose two measures to answer whether information about sentencing disparity reduces respondents' willingness to apply to a court. And (if so,) are the respondents more likely to search for alternatives to the formal judicial system? Since these questions ask about actual (intended) behavior, they provide more convincing measures of real life consequences of the treatment information than the declared level of trust.

To understand whether providing information about sentencing disparity reduces the willingness to apply to the court, I cooperated with an NGO (*vasevyzivne.cz*), which assists single-parents in filing lawsuits against a spouse who is not paying alimony. In the experiment, I briefly explain a problem of a typical client of the NGO, i.e., a single mother who is considering whether to apply to the court to sue for alimony or not. Applying to the court is potentially beneficial, but it also may lead to high cost, both in terms of money and time and no benefits. I then asked the respondents whether they would recommend her – a typical client of the NGO – to apply to the court or not. Before their answers, I informed them that the NGO might use their advice as material in similar cases. Presumably, the belief that their responses will potentially serve as a guideline for other people in actual problems increases the cost of an ill-concerned answer.

[12]For example, the government (the Minister of Justice) plays a role in appointing new judges.

Should information about sentencing disparity discourage respondents from applying to the judicial system, it seems reasonable to expect that they may be interested in substitute to the judicial system. In a similar vein, Acemoglu et al. (2020) document the substitutability between formal and informal courts motivated by perception of the poor performance of the formal courts. As the next task, I explained that in some cases it is possible to rely on alternative dispute resolution instead of the judicial system. I then offered a free booklet with basic information about alternative dispute resolution. The respondents first provided an indicative answer of their interest, and if it was affirmative, they were asked to provide their email addresses to have the booklet sent. The two-step procedure was conducted for two reasons. First, it evaded legal concerns about using their email addresses for different purposes than inviting them to conduct the survey. Second, it imposed a small but positive cost on the action.

Policy Preferences New information about the performance of public institutions' work may change policy preferences and evoke public reactions (e.g. petitions and political protests). To measure this effect, I collect two outcomes regarding the policy preferences. First, I asked respondents to imagine that they were the prime minister of the Czech Republic and gave them a list of four policy issues that they were supposed to rank according to the perceived priority. The most pressing issue was supposed to be ranked as the top priority, the second most pressing as the second priority, etc. The four policy issues were: (i) fairness of the judicial system; (ii) sufficient highway infrastructure; (iii) high-quality teachers in the education system; and (iv) the safety situation in the Czech Republic.

Second, I elicited respondents' willingness to become actively involved in policy debate. Subjects were presented with an extract of a petition inviting politicians (members of the Committee on Constitutional and Legal Affairs, Chamber of Deputies, Parliament of the Czech Republic) to suggest specifying sentencing principles. Respondents were asked to indicate their interest in reading the full text of the petition and signing it. If interested, the respondents were referred to a Google forms website with the full text of the petition. The text highlights the importance and far-reaching consequences of sentencing decisions in one's life and suggests that it may be beneficial to have a manual that would lead to more consistent sentencing. Importantly, the petition was explicit that the manual would be designed to assist judges in their sentencing decisions but would not in any way undermine their independence and discretion. If interested in signing the

petition, the respondents could have left their email address to have the signature sheet sent.[13] I collected individual declaratory answers regarding their interest in reading and signing the petition. Once respondents left the Behavio website and opened the petition, I could not observe responses at the individual level. However, since individuals in the treatment and the active control groups were referred to different forms of the identical petition, I observed the number of email addresses left for the control and the treatment group separately.

3.2.4 Randomization

The groups are balanced on both observed demographic characteristics and their prior attitudes towards the judicial system. Roughly 19% (21% in the control group) of the respondents in the treatment group reported to be single, 18% (17%) live in cohabitation, 43% (44%) are married. An additional 16% (15%) reported they are currently divorced, and only 4% (3%) are widowed. The reported marital status reflects their status at the time of the survey experiment but not their history; for example, respondents classified as married could have experienced a divorce before. Slightly more than 70% of respondents have at least one child, and the average number of children is 1.43 in the control group and 1.49 in the treatment group. In both groups, there are fewer male than female respondents (46.7% in the treatment group and 48.1% in the control group). For more details, see Table 3.4 in the appendix.

3.3 Results

3.3.1 Prior Attitude Towards the Judicial System

The majority of respondents (91.4%) agree that sentencing decisions are sensitive to the personality of a judge and that, depending on the judge assigned, similar cases can be sentenced differently. The measure of the general approval of the judicial system is less unequivocal. 52.4% agree that the judicial system works well, while 47.6% disagree. The immediate implication is that many respondents (46.7%) tend to approve of the judicial system, despite the perceived sentencing disparity. Figure 3.7 in the appendix shows the aggregate levels of responses.

[13]The respondents were informed that the petition would be filed once there are at least 1000 signatures.

Approval of the judicial system varies with the level of experience with the judicial system. More experienced respondents tend to hold more negative prior beliefs. With regard to the 10% of respondents who report the highest level of experience, the majority of them (60%) strongly agree that, depending on the judge assigned, similar cases can be sentenced differently and they do not consider the judicial system working well. Conversely. the less experienced respondents consider the judicial system working rather well and do not view the sentencing disparity as that extreme, although they still admit it may exist. Figure 3.8 in the appendix shows that more experienced respondents differ in their prior attitude towards the judicial system. The Wilcoxon rank-sum test rejects the equality of attitudes (p-value = 0.000) between the most experienced and the three levels of less experienced respondents taken together.

3.3.2 Experimental Evidence

Institutional Trust

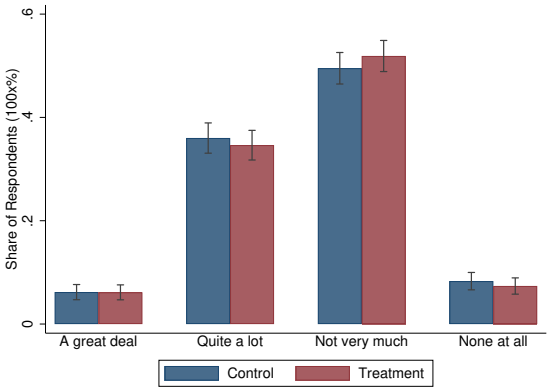
The most trusted institution is the police (56% of respondents report one of the two highest levels of trust), followed by the judicial system (42%) and the public broadcasting service (25%). The national government enjoys the lowest level of trust (21%). The results are consistent with international surveys. For example, according to Eurobarometer (2018), similar shares of Czech respondents tend to trust in the police (63%), the judicial system (43%), and the national government (28%).[14] Overall, my results, while less optimistic, tend to resemble those from Eurobarometer (2018).

A high prior level of approval of the judicial system positively correlates with the level of declared trust in all four institutions. Respondent who view the judicial system as working well is ca. 45 percentage points more likely to report a high level of trust in the judicial system. More interestingly, they are also ca. 25 percentage points more likely to trust in the police, ca. 12 percentage points in the government, and ca. 10 percentage points in the public broadcasting service. All the effects are statistically significant. The closer the institution to the judicial system is, the larger the effect is. Table 3.5 in the appendix summarizes the full results.

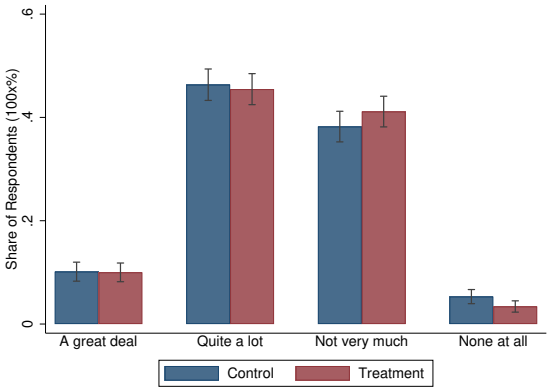
To test the effect of information about sentencing disparity on institutional trust, I conduct four exercises that all suggest that the treatment does not have an effect. I start by collapsing the first two possible answers: *a great deal* of trust; and *quite of a lot* of

[14]Eurobarometer (2018) does not include institutional trust in the public broadcasting service.

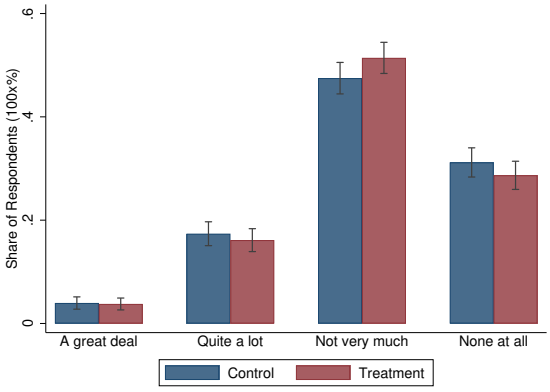
Figure 3.2: Levels of Institutional Trust



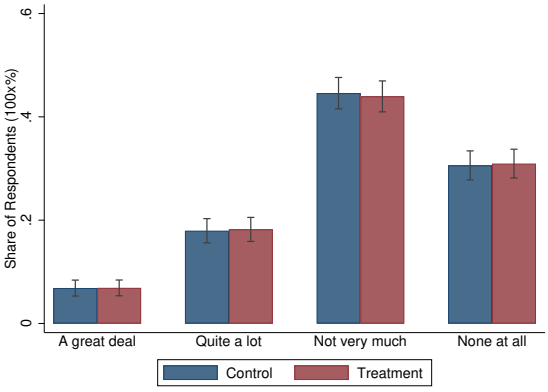
(a) Trust in Judicial System



(b) Trust in Police



(c) Trust in Government



(d) Trust in Public Broadcasting

Notes: Declared level of institutional trust by treatment status: (a) in the judicial system; (b) in the police; (c) in the government; and (d) in public broadcasting service. For each institution, respondents were asked to choose from four levels of trust: (i) a great deal; (ii) quite a lot; (iii) not very much; (iv) none at all.

trust into one category denoted as $High\,Trust$. For each institution j, I: (i) estimate univariate OLS regressions

$$High\,Trust^{j} = \alpha + \beta\ Treatment + \varepsilon \tag{3.1}$$

(ii) estimate OLS regressions with a set of dummies capturing individual prior attitude towards the judicial system and basic demographic characteristics (income, age, level of education, gender, the number of children, and heterogeneity of the treatment effect by mother-status and by prior approval of the judicial system); (iii) simulate the exact p-value for the sharp null hypothesis derived from the potential outcome framework (Athey and Imbens, 2017) and then test

$$High\,Trust_i(0) = High\,Trust_i(1) \quad \forall i = 1, \ldots, N. \tag{3.2}$$

Finally, since the levels of institutional trust represent an ordinal scale, I assign a rank (1,2,3,4) to these categories and apply the Wilcoxon rank-sum test.

To provide results that are robust against the limited attention of respondents, I work with two samples: (i) a *baseline* sample that contains all observations at the level of randomization; and (ii) a sample that drops observations for 10% of respondents who spent the least time on the treatment and control slides, respectively. That limits the number of observations to 2,168. Additionally, I also discard 10% of respondents who spent the least time on the slide with the institutional trust task, reaching 2,008 observations (83% of the initial dataset). I refer to this sample as the *restricted* sample. No effect in the *restricted* sample is credible evidence that the lack of effect is not caused by respondents' limited attention.

Table 3.1 summarizes the null results on institutional trust. Panel A shows results from the OLS regressions run on the *baseline* sample, and Panel B adds results from the *restricted* sample. The first column for each institution provides strong evidence that the average treatment effect is economically and statistically insignificant. With the exception of the public broadcasting service in the *baseline* sample, all point estimates are, in absolute value, safely less than 2 percentage points. Providing information about sentencing disparities among judges thus did not lead to institutional distrust. Similarly, the exact p-value test and the Wilcoxon rank-sum test presented in Panel C for the full sample indicate the null average treatment effect.

Table 3.1: Treatment Effect on Declared Institutional Trust

Panel A: Baseline Sample

	Judicial System		Police		Government		Broadcasting	
Treatment	-0.013	-0.007	-0.013	-0.018	-0.008	-0.016	-0.032	-0.009
	(0.032)	(0.032)	(0.039)	(0.025)	(0.030)	(0.021)	(0.033)	(0.022)
Treat. x Mother		-0.059*		-0.050		-0.034		0.011
		(0.035)		(0.040)		(0.034)		(0.034)
Prior Attitude	No	Yes	No	Yes	No	Yes	No	Yes
Demo. Char.	No	Yes	No	Yes	No	Yes	No	Yes
N	2 410	2 407	2 410	2 407	2 410	2 407	2 410	2 407

Panel B: Restricted Sample

	Judicial System		Police		Government		Broadcasting	
Treatment	-0.010	-0.008	-0.013	-0.025	-0.013	-0.006	0.003	-0.040
	(0.022)	(0.035)	(0.022)	(0.042)	(0.028)	(0.033)	(0.019)	(0.036)
Treat. x Mother		-0.057		-0.042		-0.054		0.022
		(0.038)		(0.043)		(0.036)		(0.037)
Prior Attitude	No	Yes	No	Yes	No	Yes	No	Yes
Demo. Char.	No	Yes	No	Yes	No	Yes	No	Yes
N	2 008	2 005	2 008	2 005	2 008	2 005	2 008	2 005

Robust standard errors in parentheses.

* (p<0.1), ** (p<0.05), *** (p<0.01)

Panel C: Exact p-value and p-value of the Wilcoxon test (Baseline sample)

	Judicial System	Police	Government	Broadcasting
Exact p-value	0.510	0.518	0.625	0.810
Wilcoxon test				
Full sample	0.652	0.760	0.509	0.702
Only mothers	0.063*	0.094	0.915	0.515

Panel A shows results from univariate and multivariate OLS regressions of *Treatment* on a dummy for high level (great deal of trust and quite a lot) of institutional trust.

Panel B shows results from univariate and multivariate OLS regressions of *Treatment* on a dummy for high level (great deal of trust and quite a lot) of institutional trust using the *restricted* sample of more attentive respondents.

Panel C presents p-values of two alternative measures; the exact p-value (Athey and Imbens, 2017) derived from 20,000 simulations and the p-value of the Wilcoxon rank-sum test that tests for the same rank of institutional trust.

The null average treatment effect might mask heterogeneity. Among mothers[15], who are arguably more sensitive to inconvenient information regarding *failure to pay alimony*, the treatment tends to marginally increase distrust. The second columns for each institution in Panel A and B report results with a focus on $Treatment\,x\,Mother$, controlling for prior beliefs and attitude towards the judicial system and demographic characteristics, such as age, gender, level of education, income, and whether a respondent is a mother. Despite a marginal statistical significance, Panel A and the Wilcoxon test in Panel C suggest treatment heterogeneity. Taking the point estimates at their face values, the share of trusting respondents among mothers decreases by 6 percentage points more than among non-mother respondents. I did not find heterogeneous treatment effect with respect to a prior level of approval of the judicial system. The treatment information did not affect the trust of those who approve of the judicial system in the same way as it did not affect those who do not approve of the judicial system. Table 3.5 in the appendix summarizes the full results.

Reliance on the Judicial System

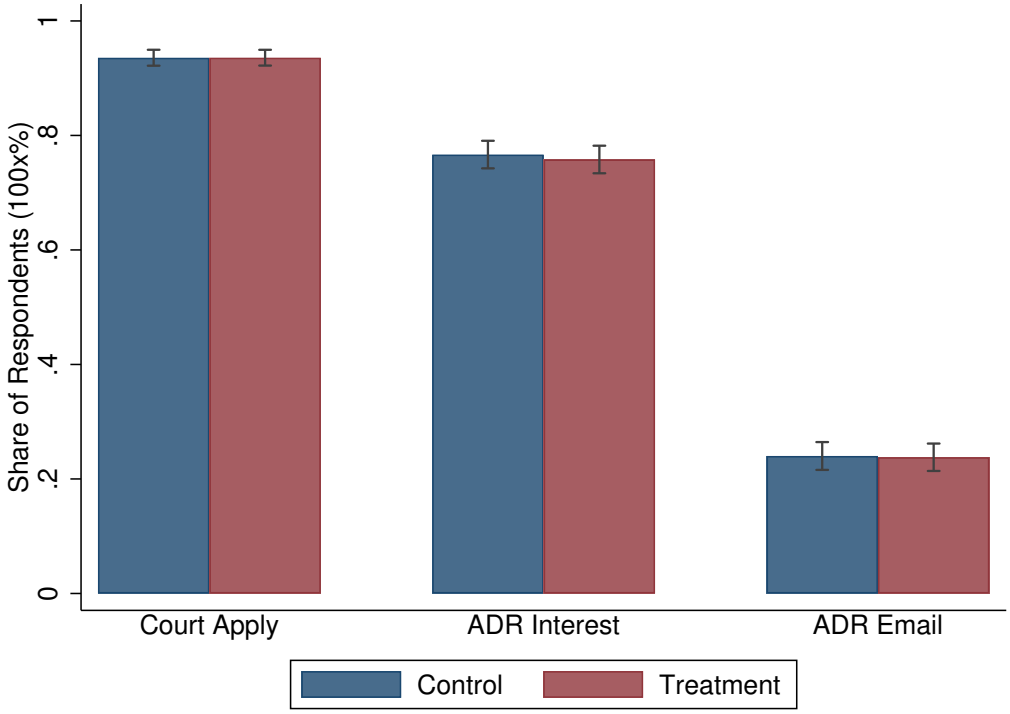
The second group of outcomes consists of a measure of respondents' intention to apply to a court, $Court\,Apply$, and two measures of their willingness to learn about alternative dispute resolution $ADR\,Interest$ and $ADR\,Mail$. In general, respondents exhibited a strong will to apply to the judicial system, as almost 93.6% of respondents recommended that a single mother apply to the court in a situation in which she hesitates. A high level of prior approval of the judicial system correlates with this decision. In particular, respondents who consider the judicial system working well, are 4 percentage points more likely to recommend that a single mother apply to the court.

If offered, a majority (76.2%) would be interested in receiving a booklet about alternative dispute resolutions. However, when the respondents were asked to provide their email addresses to have the booklet sent, only 23.9% of all respondents and 31.4% of those who declared their interest did so. See Figure 3.3. Better educated and more experienced respondents are more likely to demand ADR, which suggests that demand for alternative dispute resolution requires a particular level of sophistication regarding the judicial system. See Table 3.6 in the appendix.

Information about sentencing disparities affected neither the (intended) reliance on

[15]A female respondent is characterized as a mother if she has at least one child, regardless of the child's age.

Figure 3.3: Reliance on Judicial System



Notes: Shares of affirmative responses by the treatment status. *Court Apply* captures whether respondents recommend that a single mother apply to the court when she hesitates. *ADR Interest* measures shares of respondents who indicate their interest in receiving information about alternative dispute resolution (ADR). *ADR Mail* shows shares of respondents who provide their email addresses to have the information about ADR sent.

the judicial systems nor the demand for alternative dispute resolution. Table 3.2 reports results from a univariate OLS regression, a multivariate regression with a focus on treatment effect on mothers, and an exact p-value. The first columns for each measure show that the average treatment effects are statistically insignificant and economically negligible. In terms of the magnitude of the point estimates, the effects are even smaller than the one on institutional trust. All point estimates of the average treatment effects are bounded between -1 and 1 percentage points. Panel B shows results estimated on a sample of respondents that spent enough time on the treatment and control slides and on the slide with the corresponding task and confirms the null results.[16]

With regard to the heterogeneity of the treatment effect, not even more sensitive groups of respondents change their behavior. In particular, mothers tend to respond to the inconvenient information about the sentencing disparity in the same manner other respondents did, i.e. not at all. Similarly, there is no heterogeneous treatment effect with respect to approval of the judicial system. Table 3.6 in the appendix shows the full results. Panel C in Table 3.2 shows the exact p-value and provides additional evidence of strong null results for all three measures.

Policy preferences

Petition I collected two measures of respondents' reactions to the petition. At the individual level, I measured respondents' interest in reading and signing the petition as an indicator that can be linked to other characteristics of respondents. At the control and treatment groups' level, I collected the number of email addresses provided. After declaring one's interest in reading and signing the petition, the respondents were referred to an external website. Once they left the platform, it is impossible to track their decisions at the individual level and match it to their characteristics. However, since the control and the treatment groups were referred to different petitions (with identical text), I can measure the number of email addresses left by respondents in the control and the treatment group.

More than 60% of respondents showed their interest in reading and signing the petition. Higher education, more experience with the judicial system, and perception of a sizeable sentencing disparity are correlated with higher likelihood of signing and reading the petition. Figure 3.4 shows the shares of respondents interested in reading and

[16]Since the overlap of the dropped samples differs across tasks, the number of observations does too.

Table 3.2: Treatment Effect on Reliance on Judicial System

Panel A: Baseline Sample

	Court Apply		ADR Interest		ADR Mail	
Treatment	0.000	0.021	-0.009	-0.027	-0.002	0.009
	(0.010)	(0.022)	(0.017)	(0.035)	(0.017)	(0.033)
Treat. x Mother		-0.018		0.049		0.030
		(0.021)		(0.036)		(0.035)
Prior Attitude	No	Yes	No	Yes	No	Yes
Demo. Char.	No	Yes	No	Yes	No	Yes
N	2 410	2 407	2 410	2 407	2 410	2 407

Panel B: Restricted Sample

	Court Apply		ADR Interest		ADR Mail	
Treatment	0.002	0.008	-0.007	-0.020	0.006	0.028
	(0.010)	(0.022)	(0.018)	(0.037)	(0.019)	(0.010)
Treat. x Mother		0.010		0.051		0.010
		(0.021)		(0.037)		(0.039)
Prior Attitude	No	Yes	No	Yes	No	Yes
Demo. Char.	No	Yes	No	Yes	No	Yes
N	2 020	2 017	2 023	2 020	2 023	2 020

Robust standard errors in parentheses

* (p<0.10), ** (p<0.05), *** (p<0.01)

Panel C: Exact p-value (Baseline sample)

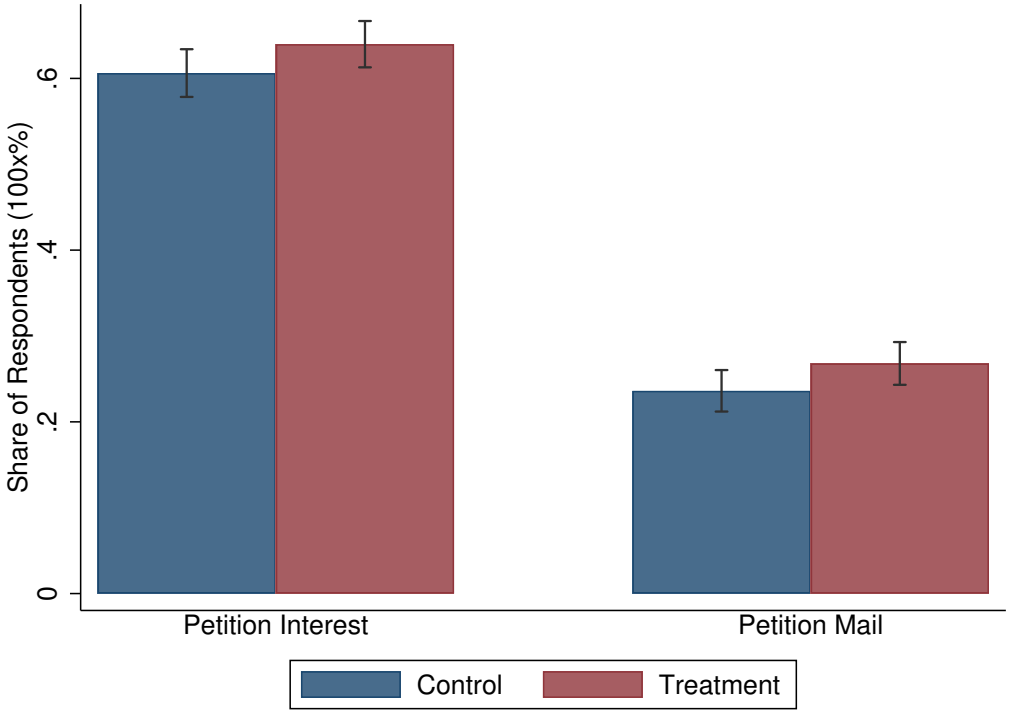
	Court Apply	ADR interest	ADR Mail
Exact p-value	0.993	0.620	0.898

Panel A shows results from univariate and multivariate OLS regressions of *Treatment* on a dummy variable for affirmative response from three tasks.

Panel B shows results from univariate and multivariate OLS regressions of *Treatment* on a dummy variable for affirmative response from three tasks using the *restricted* sample of respondents attentive respondents.

Panel C shows the exact p-value (Athey and Imbens, 2017) derived from 20,000 simulations.

Figure 3.4: Interest in Reading and Signing Petition.

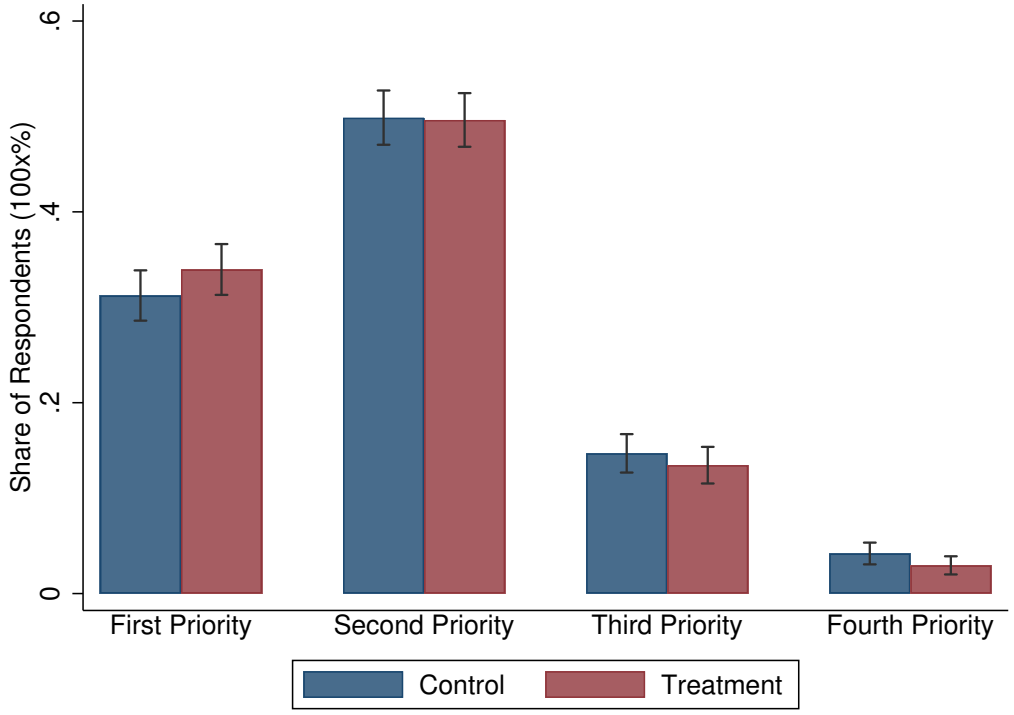


Notes: Share of respondents who declare their interest in reading and signing petition. *Petition Interest* measures shares of respondents of the *baseline* sample who indicate their interest in reading the petition. *Petition Mail* measures shares of respondents of all respondents invited in the survey who provided me with their email addresses to have the petition sent.

signing the petition and those who provided their email addresses by the control and the treatment groups. The share was higher in the treatment group by 3.3 percentage points (5.8%). Once the respondents were asked to provide their email addresses, the share of affirmative action dropped to 25%. The gap between the treatment and the control groups remains almost constant at 3.2 percentage points (13.6%).

Formal tests suggest that the average treatment effects on: (i) respondents' interest in reading and signing the petition; and (ii) providing an email address are marginally significant. The former effect becomes slightly larger once I drop observations of the least attentive respondents. See Panel B in Table 3.3. The average treatment effect, however, masks sizeable heterogeneity. The second column of Panel A shows that the effect is more pronounced among mothers who are more likely to act on that information. In particular, the inconvenient information about the sentencing disparity in cases of *failure to pay alimony* increases the likelihood that a mother will be interested in reading and signing the petition by 10.7 percentage points compared to mothers in the control group.

Figure 3.5: Rank of Fairness of Judicial System as Priority



Notes: Share of respondents who rank fairness of the judicial system as the first, second, third, and the fourth priority by the control and treatment groups. The remaining issues to be ranked were sufficient highway infrastructure, safety in the Czech Republic, and high-quality teachers in the education system.

The effect is even larger on the *restricted* sample. Full results in Table 3.7 show that there is no heterogeneity in the treatment effect with respect to the level of approval.

Policy preferences The respondents view the fairness of the judicial system as a relevant policy issue. A third (32.6%) of them ranked fairness of the judicial system as the top priority and an additional 50% as the second most important priority. While the perception of fairness of the judicial system is likely affected by the survey experiment itself (e.g., through the experimenter demand effect) and thus it is barely generalizable, it is a good signal of the relevance of the issue. Figure 3.5 shows shares of respondents who ranked fairness of the judicial system as the first, the second, the third, and the fourth priority by both the treatment and the control groups. The figure suggests that the ranking among the treated respondents is slightly shifted towards the higher priority compared to the control group.

The share of respondents who would address the fairness of the judicial system as the top priority in the treatment group is 34%, while in the control group, it is 31.2%.

Table 3.3: Treatment Effect on Political Preferences

Panel A: Baseline Sample

	Petition		Petition Mail	Top Priority JS	
Treatment	0.033*	0.030	0.032*	0.027	0.073*
	(0.019)	(0.040)	(0.017)	(0.019)	(0.038)
Treat. x Mother		0.077*			-0.008
		(0.040)			(0.039)
Prior Attitude	No	Yes	No	No	Yes
Demo. Char.	No	Yes	No	No	Yes
N	2 410	2 407	2 410	2 410	2 407

Panel B: Restricted Sample

	Petition		Petition Mail	Top Priority JS	
Treatment	0.047**	0.042		0.034*	0.070*
	(0.021)	(0.042)		(0.020)	(0.041)
Treat. x Mother		0.100**			-0.005
		(0.043)			(0.042)
Prior Attitude	No	Yes		No	Yes
Demo. Char.	No	Yes		No	Yes
N	2 037	2 034		2 018	2 015

Robust standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Panel C: Exact p-value (Baseline sample)

	Petition	Petition Mail	Top Priority JS
Exact p-value	0.086*	0.079*	0.155

Panel D: Rank Judicial System - Wilcoxon test

	Baseline sample	Restricted sample
Wilcoxon test	0.060*	0.041**

Panel A and B show results from OLS regressions of *Treatment* on dummy variables that measure whether one (i) is interested in reading a petition; (ii) provide email address to have the petition sent; (iii) ranks fairness of judicial system as the top or second priority.

Panel B is estimated on a sample without the least attentive respondents.

Panel C shows the exact p-value (Athey and Imbens, 2017) derived from 20,000 simulations.

The difference is not statistically significant. Once one looks at the *restricted* sample of more attentive respondents, it becomes marginally significant. See table 3.3. The multivariate regressions reveal a sizeable heterogeneity with respect to previous experience with the judicial system. The overall effect aggregates a positive effect of less experienced respondents and negative effect of the more experienced. See the full results in Table 3.7. To formally test the differences in the rankings of the policy issues, I rely on the Wilcoxon rank-sum test. Panel D in Table 3.3 shows that the ranks marginally differ, especially in the restricted sample. That confirms the apparent shift in the ranking of the judicial system among the treatment group. There seems to be no heterogeneity effect with respect to mother-status and the level of approval.

3.4 Discussion

3.4.1 Interpretation of the Treatment Effect

An important concern in studying trust and trustworthiness based on information treatment is whether respondents trust the provided information. At the beginning of the experiment, respondents were informed that the provided information is truthful and based on data from the Ministry of Justice. This may evoke a tension between implicitly asking respondents to trust the data provided by the Ministry of Justice and, at the same time, asking them whether they trust the judicial system. To understand the degree of the potential risk, I elicited the perceived credibility of the information. To avoid influencing the experiment, the question regarding the credibility of the information came at the end of the experiment. 87% of respondents declared that they view the data as credible and only less than 1% of them selected an extreme choice of *definitely not credible*. The shares are virtually identical in both groups and the results are not systematically affected by respondents' mistrust in the information treatment. See figure 3.9 in the appendix.

For the proper interpretation of the results, it is important how respondents understand and interpret the treatment information. The core of the treatment is to provide details on varying levels of sentencing disparity, i.e., the second moment of the distribution of sentencing decisions. This is a novel feature. Most of the information and survey experiments existing in the literature exogenously vary beliefs about the first moment of a relevant distribution, e.g., a probability of audit, a share of high-skilled immigrants. Varying the signals about sentencing disparity is challenging, as it may be more compli-

cated for the general public to understand the information and interpret it correctly.

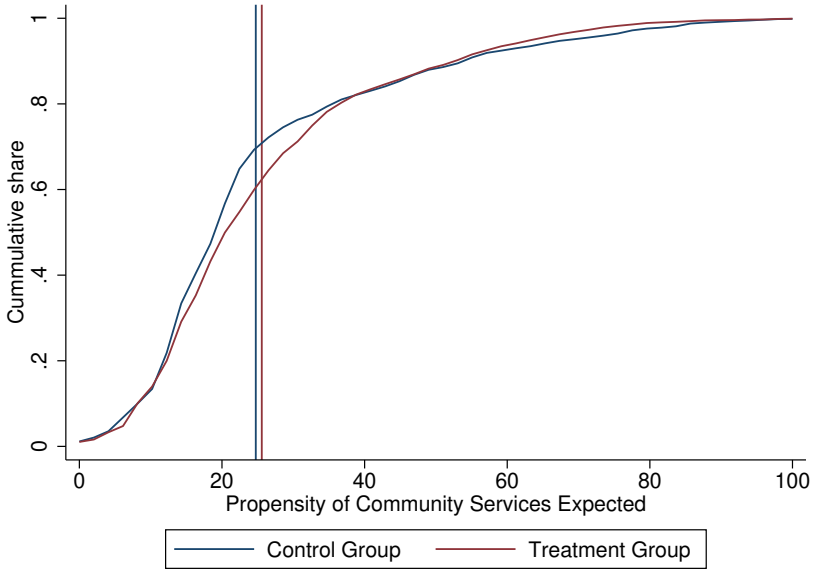
Additionally, one varies other properties (e.g. maximum, minimum) of the data used to derive the information treatment. Conveniently, in the two courts used in the control and the treatment groups, the average shares of convicted offenders sentenced to community service for *failure to pay alimony*, i.e. the first moment of the distribution, were numerically identical (16%). Nevertheless, it is still plausible that the provided information affects respondents' perception of the propensity to sentence to community service systematically differently in the control and the treatment groups.

To understand this threat, I elicited the respondents' expectation regarding the average propensity to be sentenced to community service, i.e., the first moment of the distribution. In particular, the respondents were asked in what percentage of cases in the Czech Republic of *failure to pay alimony* is a convicted offender sentenced to community service. On average, respondents in both groups overestimate the actual shares. While the national average corresponds to the presented cases, i.e. 16%, the respondents in the control group expected 24.7% and in the treatment groups 25.6%. The difference between the groups is not statistically significant (p-value = 0.255). Figure 3.6 shows empirical cumulative distribution functions of the expected share of cases sentenced to community service and suggests that most of their estimates, in both groups, are concentrated between 10% and 30%. The fact that the cumulative distributions functions resemble each other, and the averages are not statistically different, suggests that the information provided does not affect the perception of the propensity of sentence to community services systematically differently in the control and the treatment groups.

3.4.2 Implications of the Results

I view three important implications of my results. First, evidence that inconvenient information did not lead to a decline in institutional trust and willingness to rely on formal institutions limits the concerns that revealing the information would be harmful from the public perspective. However, the personal incentives of public officers who decide whether to publish the information or not may still prevent publishing. If a public officer suspects that the information may harm him, his reputation, and his future in office, he may, in order to keep the information confidential, argue that if the information were public, it would cause distrust with a high (social) cost. My results imply that using that argument sounds more like a pretext than a real concern. Instead, the general public is

Figure 3.6: Expected Share of Cases Sentenced to Community Service



Notes: The figure shows empirical cumulative distribution function of respondents' estimates of the share of cases that are typically sentenced to community service at the national level for both the control and the treatment group.

likely to demand policy changes, which may indeed jeopardize position of the incumbent public officer.

Second, the observed heterogeneity shows the extent to which a particular group of citizens can drive the reaction to inconvenient information. While the idea is not new, I provide empirical estimates of such an effect. It suggests that even issues as worrying as sentencing disparities may remain overlooked and ignored as long as the information is not provided or available to a particular group of citizens. This is likely to hold more generally in many other policy issues, such as inefficiency of public procurement contracts. The heterogeneity also implies that publishing information about sentencing disparities for more (all) offenses would likely lead to a sizeable increase in the overall effect, as each of the offenses is likely to trigger additional groups of citizens based on their sensitivity to the particular topic and offense. From that perspective, the effect estimated in this study would represent a lower bar of the effect.

Third, my results question the information value of standard measures of institutional trust. Even though the information increases the likelihood of signing a petition and demanding a change of the current system, suggesting dissatisfaction with situation, the information however did not pass-through to the standard measures of trust. Suppose

policymakers and international organizations identify social issues and consequently build policies and recommendations based on survey measures of institutional trust alone. In that case, it raises the possibility of missing an important feature of citizens' preferences and dissatisfaction with formal institutions.

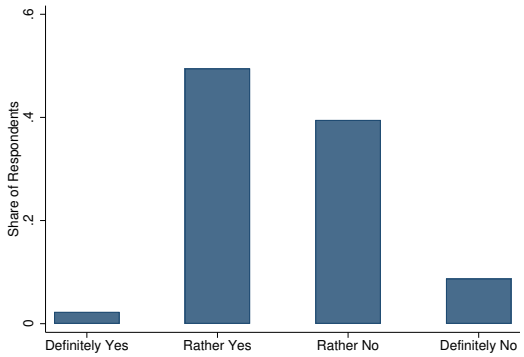
3.5 Concluding Remarks

Publishing inconvenient information about the performance of public institutions in an environment where market mechanisms cannot operate raises the question of how citizens would respond. I focus on a particular case of sentencing disparities that undermine the principles of a clear, stable, and predictable application of law, and consequently, equality before the law. The results suggest that inconvenient information about sentencing disparities does not lead to distrust and avoidance of the formal judicial systems. Instead, respondents exposed to the information found the fairness of the judicial system to be a more important policy issue and were more likely to sign a petition that calls on politicians to address the issue. Additionally, there seems to be a sizeable heterogeneity in the treatment effect. A personal connection to the offense seems more important for the effect of the information than a prior belief regarding the performance of that institution.

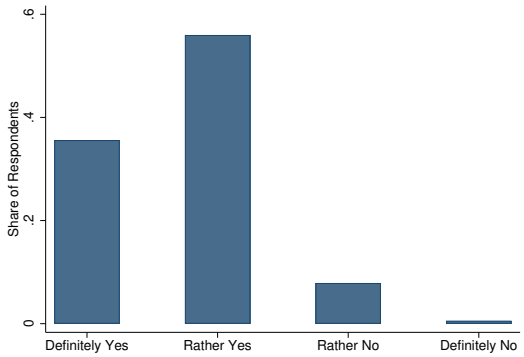
The results imply that the concern inconvenient information would lead to losing faith in the formal institutions does not seem to be of first-order importance. However, the external validity of the results is limited by several considerations. First, the information provided in the experiment is one-time, brief, and isolated. If the same information makes it in the headline news and becomes widespread, discussed in the spotlight, among peers, colleagues, and politicians, it would be more difficult to ignore the information. As a result, my results likely underestimate the potential effects of inconvenient information. However, even if the effects were of a considerably larger magnitude, my results are arguably still informative about the degree of the impact on the particular behavior. This is consistent with the aftermath of George Floyd's death: even though some Americans prefer defunding or abolishing the police, the majority support reform (Vaughn et al., 2021). Second, evidence about heterogeneity implies that a different offense is likely to trigger different citizens. Therefore, publishing all information about all offenses would likely increase the effect as well. Finally, since I focus on a particular case of the judicial system, it is not clear how the results generalize in different domains of public institutions.

3.6 Appendix

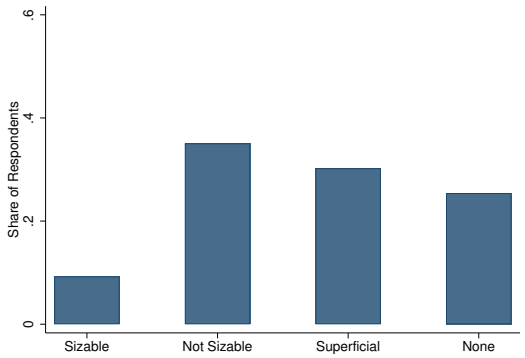
Figure 3.7: Prior Attitude Towards Judicial System



(a) Prior Approval of the JS



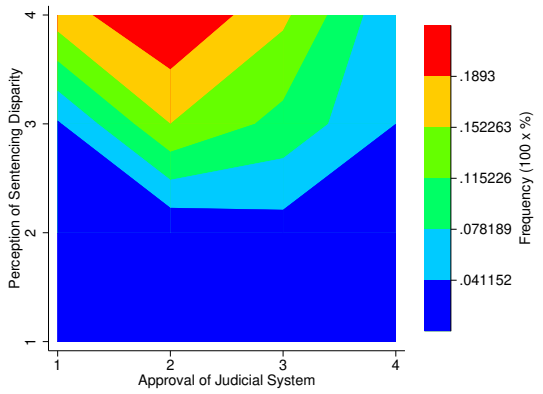
(b) Perception of Sentencing Disparity



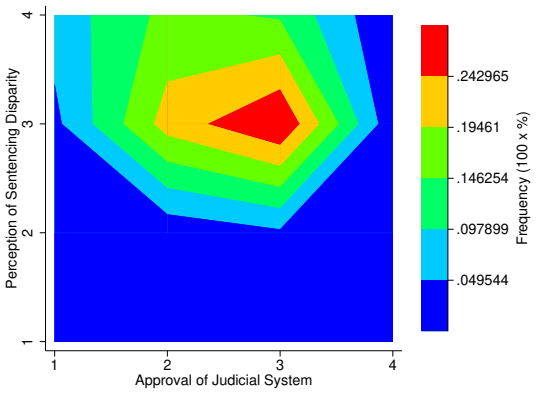
(c) Prior Experience with the JS

(a) The judicial system in the Czech Republic works well. (b) Judges regularly differ in sentencing decisions in similar cases. (c) Considering how often you or people you know well come into contact with the judicial system, how experienced do you think you are?

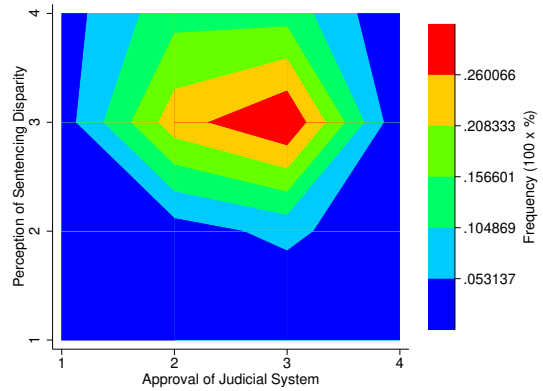
Figure 3.8: Attitude Towards Judicial System by Experience



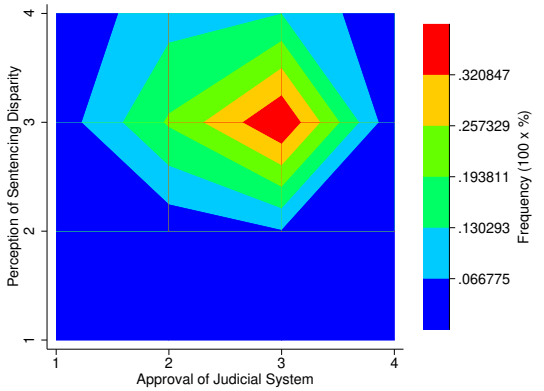
(a) Sizable Experience With the JS



(b) Not Sizeable Experience with the JS



(c) Superficial Experience with the JS



(d) No Experience with the JS

Notes: Level of approval of the judicial system (x-axis) and perception of sentencing disparity (y-axis) by groups of respondents with different levels of experience with the judicial system.

Table 3.4: Mean Characteristics of Treatment and Control Groups

Variable	Mean Control	Mean Treatment	t-test (p-value)
Measures of Prior Attitude			
Approval of Judicial System			
Definitely yes	0.026	0.020	0.345
Rather yes	0.486	0.499	0.571
Rather no	0.393	0.397	0.870
Definitely no	0.094	0.085	0.437
Perception of Sentencing Disparity			
Definitely yes	0.367	0.338	0.161
Rather yes	0.551	0.572	0.311
Rather no	0.077	0.084	0.532
Definitely no	0.006	0.006	0.967
Experience with the Judicial System			
Sizable	0.089	0.098	0.479
Not sizeable	0.347	0.343	0.857
Superficial	0.307	0.307	0.964
None	0.257	0.253	0.820
Demographic Characteristics			
Education			
University	0.247	0.251	0.823
Highschool	0.697	0.694	0.849
Elementary	0.055	0.055	0.968
Marital Status			
Single	0.210	0.187	0.116
Cohabitation	0.173	0.183	0.523
Married	0.438	0.426	0.567
Divorced	0.151	0.164	0.413
Widowed	0.028	0.041	0.049
Male	0.481	0.467	0.538
Age	44.61	45.08	0.435
At least 1 child	0.714	0.723	0.622
Number of children	1.425	1.493	0.193
N	1,071	1,097	

Table 3.5: Treatment Effect on Declared Institutional Trust (Full Results)

	Judicial System		Police		Government		Broadcasting	
Treatment	-0.007	-0.008	-0.018	-0.025	-0.016	-0.006	-0.032	-0.040
	(0.032)	(0.035)	(0.039)	(0.042)	(0.030)	(0.033)	(0.033)	(0.036)
Treat.x Mother	-0.059*	-0.057	-0.050	-0.042	-0.034	-0.054	0.011	0.022
	(0.035)	(0.038)	(0.040)	(0.043)	(0.034)	(0.036)	(0.034)	(0.037)
Mother	0.031	0.007	0.005	-0.027	-0.027	-0.038	0.022	-0.016
	(0.042)	(0.047)	(0.047)	(0.053)	(0.039)	(0.043)	(0.045)	(0.050)
High SD	-0.100***	-0.098***	-0.050*	-0.058	0.013**	0.010	-0.049	-0.072
	(0.030)	(0.033)	(0.031)	(0.034)	(0.029)	(0.032)	(0.033)	(0.036)
High Approval	0.469***	0.460***	0.268***	0.240***	0.134***	0.117***	0.116***	0.096***
	(0.025)	(0.027)	(0.028)	(0.031)	(0.023)	(0.026)	(0.025)	(0.027)
High Experience	-0.065***	-0.064**	-0.056**	-0.044***	-0.043*	-0.039	-0.049**	-0.075***
	(0.024)	(0.027)	(0.028)	(0.031)	(0.024)	(0.026)	(0.025)	(0.026)
Treat. x High Appr.	0.007	0.020	0.072*	0.087**	-0.005	-0.005	-0.032	-0.024
	(0.034)	(0.037)	(0.039)	(0.043)	(0.032)	(0.036)	(0.035)	(0.037)
Treat. x High Exper.	0.032	0.024	-0.031	-0.035	0.043	0.036	0.099***	0.107***
	(0.034)	(0.037)	(0.039)	(0.043)	(0.033)	(0.036)	(0.035)	(0.037)
Male	0.037	0.016	0.060*	0.031	0.027	0.012	0.086**	0.040
	(0.032)	(0.037)	(0.036)	(0.041)	(0.029)	(0.033)	(0.037)	(0.042)
Age	-0.010**	-0.010**	-0.011**	-0.013**	-0.011**	-0.012**	0.007	0.010*
	(0.004)	(0.005)	(0.005)	(0.006)	(0.005)	(0.005)	(0.005)	(0.005)
Age Sq	0.000	0.000	0.000*	0.000*	0.000***	0.000***	-0.000**	-0.000**
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
University Educ.	0.092***	0.084***	-0.014	-0.021	-0.019	-0.019	0.098***	0.101***
	(0.020)	(0.022)	(0.023)	(0.026)	(0.019)	(0.021)	(0.022)	(0.024)
Income (1000 CZK)	-0.000	-0.000	0.001	0.001	-0.001***	-0.002**	0.002*	0.002**
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Child Dummy	0.014	-0.015	-0.012	-0.009	0.041	0.064*	-0.028	-0.044
	(0.034)	(0.037)	(0.039)	(0.043)	(0.033)	(0.035)	(0.036)	(0.039)
Number Children	0.016	0.019*	0.020	0.022	0.005	-0.000	-0.023**	-0.019*
	(0.011)	(0.011)	(0.013)	(0.014)	(0.011)	(0.012)	(0.010)	(0.010)
Constant	0.568***	0.612***	0.686***	0.791***	0.252***	0.274***	0.114	0.107
	(0.094)	(0.105)	(0.106)	(0.119)	(0.092)	(0.104)	(0.098)	(0.110)
Observations	2407	2005	2407	2005	2407	2005	2407	2005
Restricted Sample	No	Yes	No	Yes	No	Yes	No	Yes

Standard errors in parentheses

* $p < 0.01$, ** $p < 0.05$, *** $p < 0.01$

Table 3.6: Treatment Effect on Reliance on Judicial System (Full Results)

	CourtApply		ADR Interest		ADR Mail	
Treatment	0.021	0.008	-0.027	-0.020	0.009	0.028
	(0.022)	(0.022)	(0.035)	(0.037)	(0.033)	(0.037)
Treat.x Mother	-0.018	0.010	0.049	0.051	0.030	0.010
	(0.021)	(0.021)	(0.036)	(0.037)	(0.035)	(0.039)
Mother	0.015	0.004	-0.017	-0.049	0.029	0.004
	(0.025)	(0.026)	(0.041)	(0.043)	(0.045)	(0.050)
High SD	-0.029**	-0.012	-0.025	-0.036	0.054	0.051
	(0.014)	(0.015)	(0.031)	(0.031)	(0.029)	(0.032)
High Approval	0.040***	0.048***	0.008	0.014	-0.002	0.005
	(0.015)	(0.015)	(0.024)	(0.026)	(0.025)	(0.027)
High Experience	-0.011	-0.000	0.086***	0.081***	0.050**	0.056**
	(0.015)	(0.015)	(0.025)	(0.026)	(0.025)	(0.028)
Treat. x High Appr.	-0.020	-0.015	0.018	0.018	-0.048	-0.059
	(0.021)	(0.021)	(0.035)	(0.037)	(0.035)	(0.039)
Treat. x High Exper.	-0.006	0.001	-0.021	-0.035	0.008	0.014
	(0.021)	(0.021)	(0.035)	(0.037)	(0.0353)	(0.039)
Male	-0.004	0.003	-0.018	-0.043	0.062*	0.039
	(0.019)	(0.020)	(0.031)	(0.033)	(0.035)	(0.0402)
Age	-0.001	-0.002	-0.003	-0.006	-0.005	-0.005
	(0.003)	(0.003)	(0.005)	(0.005)	(0.005)	(0.005)
Age Sq	0.000	0.000	0.000	0.000	0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
University Educ.	0.006	0.006	0.117***	0.100***	0.071***	0.076***
	(0.012)	(0.011)	(0.019)	(0.020)	(0.022)	(0.025)
Income (1000 CZK)	0.001**	0.001**	0.003***	0.002***	0.001	0.001
	(0.000)	(0.000)	(0.001)	(0.001)	(0.001)	(0.001)
Child Dummy	0.002	0.018	-0.055	-0.035	-0.103***	-0.086**
	(0.021)	(0.021)	(0.035)	(0.037)	(0.036)	(0.041)
Number Children	-0.001	-0.004	-0.001	-0.001	0.024**	0.021*
	(0.007)	(0.007)	(0.011)	(0.012)	(0.012)	(0.013)
Constant	0.946***	0.959***	0.770***	0.903***	0.247***	0.272**
	(0.057)	(0.056)	(0.100)	(0.106)	(0.094)	(0.109)
Observations	2394	2010	2407	2020	2407	2020
Restricted Sample	No	Yes	No	Yes	No	Yes

Robust standard errors in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

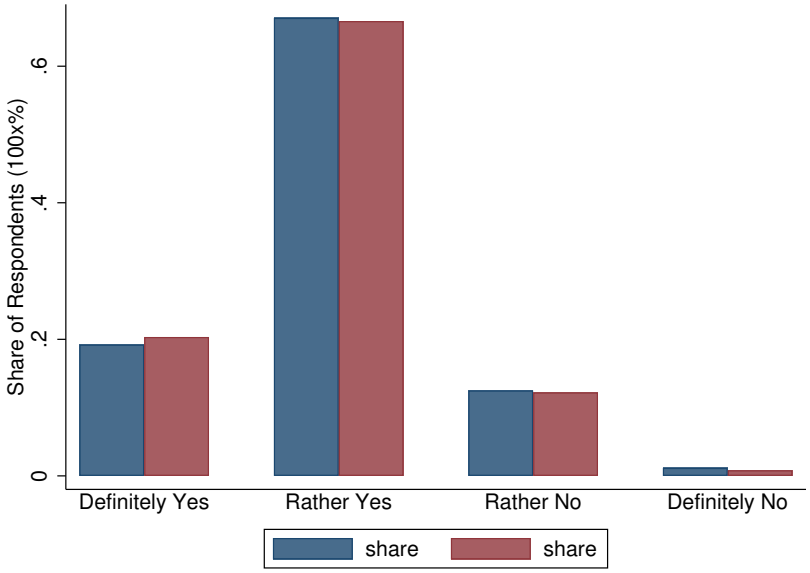
Table 3.7: Treatment Effect on Policy Preferences (Full Results)

	Petition		Top Priority JS	
Treatment	0.030	0.042	0.073*	0.070*
	(0.040)	(0.042)	(0.038)	(0.041)
Treat.x Mother	0.077*	0.100**	-0.008	-0.005
	(0.040)	(0.043)	(0.039)	(0.042)
Mother	0.010	-0.024	0.036	0.056
	(0.050)	(0.054)	(0.048)	(0.053)
High SD	0.075**	0.093**	-0.032	-0.024
	(0.036)	(0.034)	(0.035)	(0.037)
High Approval	0.020	0.015	-0.059**	-0.065**
	(0.029)	(0.031)	(0.027)	(0.030)
High Experience	0.076***	0.074**	0.072***	0.058*
	(0.029)	(0.031)	(0.027)	(0.030)
Treat. x High Appr.	-0.058	-0.054	0.003	0.005
	(0.040)	(0.043)	(0.039)	(0.042)
Treat. x High Exper.	0.002	-0.017	-0.097**	-0.082*
	(0.040)	(0.043)	(0.039)	(0.043)
Male	0.011	0.001	-0.017	0.005
	(0.038)	(0.042)	(0.037)	(0.042)
Age	0.000	-0.001	0.002	0.000
	(0.005)	(0.006)	(0.005)	(0.006)
Age Sq	0.000	0.000	-0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)
University Educ.	0.055**	0.048*	-0.018	-0.010
	(0.024)	(0.026)	(0.023)	(0.025)
Income (1000 CZK)	0.000	0.000	0.001	0.001
	(0.001)	(0.001)	(0.001)	(0.001)
Child Dummy	-0.049	-0.008	-0.096**	-0.077*
	(0.040)	(0.044)	(0.039)	(0.044)
Number Children	0.012	0.003	0.005	0.003
	(0.012)	(0.013)	(0.013)	(0.014)
Constant	0.426***	0.480***	0.309***	0.313***
	(0.112)	(0.125)	(0.107)	(0.120)
Observations	2407	2034	2407	2015
Restricted Sample	No	Yes	No	Yes

Robust standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Figure 3.9: Perception of Credibility of Information Provided



Notes: The figure shows shares of respondents classified by how credible they perceive information provided by the experimenter.

Script

Throughout the experiment, the respondents were shown 14 or 15 slides, depending on their choices. Furthermore, at the end of the session, the respondents could choose to go to a website containing a petition to sign. The survey experiment consists of three blocks: (i) prior-treatment questions, (ii) treatment slides, and (iii) post-treatment questions.

Prior-treatment Questions

Prior to the treatment, respondents are asked three questions to elicit their prior beliefs regarding the sentencing disparity, their experience with the judicial system, and their approval of the judicial system.

Slide 1. You are about to participate in a survey about courts in the Czech Republic. All information provided is truthful and based on data from the Ministry of Justice.

Slide 2. We are going to show you three statements and ask you how much you agree with these statements.

Slide 3. Sentencing decisions depend on the particular judge assigned to the case. Judges regularly differ in sentencing decisions in similar cases.

- Definitely yes
- Rather yes
- Rather no
- Definitely no

Slide 4. The judicial system in the Czech Republic works well.

- Definitely yes
- Rather yes
- Rather no
- Definitely no

Slide 5. Considering how often you or people you know well come into contact with the judicial system, how experienced do you think you are?

- Sizable
- Not sizeable
- Superficial
- None

Treatment

In this section of the survey, I show different information to the control group and to the treatment group. Both the control and the treatment slides consist of a figure of judges showing shares of cases sentenced to community service and a brief explanation of what the figure represents. Figure **??** (**??**) shows the control (treatment) slide.

Slide 6. One of the most frequent crimes in the Czech Republic is *failure to pay alimony*. In the last 3 years, courts in the Czech Republic have sentenced more than 13,000 cases. The punishment can be a suspended sentence, incarceration, community service, and/or a fine.
Now, we will show you sentencing decisions of judges at a regional court in the Czech Republic.

Slide (Control Group). At this regional court, 16% of the convicted are sentenced to community service as their main punishment.
Judges sentence very similarly.
Regardless of which judge is assigned to the case, the offender has very similar probability of being sentenced to community service.

Slide (Treatment Group). At this regional court, 16% of the convicted are sentenced to community service as their main punishment.
Judges sentence differently.
Judge C sentences a third of his/her cases to community service.
Other judges sentence less than 10% cases to community service, instead they choose different types of punishments.
Cases are assigned to judges at random: an offender assigned to judge C has three times higher probability of being sentenced to community service compared to a situation in which he was assigned to a different judge.

Post-treatment Questions

Slide 7. The figure showed the situation from one regional court. Based on your opinion, what is the situation in the Czech Republic? In what % of cases are offenders sentenced to community service as the main punishment for *failure to pay alimony*?

Slide 8. How much do you trust the following institutions ... { the judicial system; the police; the government; public broadcasting }

- A great deal
- Quite a lot
- Not very much
- None at all

Slide 9. Suppose you are the prime minister of the Czech Republic. Rank the following issues according to the priorities you would approach them.

- Fair judicial system
- Sufficient highway infrastructure
- High-quality teachers in the education system
- Safety situation in the Czech Republic

Slide 10. In some cases, it is possible to substitute the formal judicial system by alternative dispute resolution (arbitration), which has several advantages compared to the judicial system.

- Want to know more
- Not interested

Slide (only if Slide 10: want to know more). If you are interested, we can send you a booklet with information about arbitration. What is your email address:

Slide 11. We would like your advice. Your response can be used by a NGO (vasevyzivne.cz) as information for its clients. Please read the following story: Jane has two kids and their father does not pay alimony. She hesitates to apply to the court, because she is not sure whether a court could help or it would be only a waste of time and energy.
What would you recommend to her?

- She should apply to the court
- There is no good reason to apply to the court

Slide 12. The last question: Would you sign a petition that invites politicians to suggest specifying sentencing principles? Such principles would assist judges in making their sentencing decision. (A preview of the petition was shown.)

- Want to read it
- Not interested at all

Slide 13. Great! That is the end of the questionnaire and very last thing (if slide 12: interested: and before we show you the petition promised), how credible do you think the information provided here is?

- A great deal
- Quite a lot
- Not very much
- Not at all

Petition

The aim of the following petition is to prompt political representatives to address the specifics of sentencing principles. The petition is addressed to members of the Committee on Constitutional and Legal Affairs of the Chamber of Deputies, Parliament of the Czech Republic and will be sent once there are at least 1000 signatures collected.
Should you be interested in signing this petition, please provide us with your email address and we will send you the signature sheet.

To members of the Committee on Constitutional and Legal Affairs of the Chamber of Deputies, Parliament of the Czech Republic
Subject: Invitation to specify sentencing principles

Sentencing decisions often crucially affect one's life. The difference between being sentenced to imprisonment and being given a suspended sentence has significant consequences for the offender, his family and friends.

A judge assigned to a case has an exclusive right to make the sentencing decision. In making such important decisions, the judge considers many circumstances related to the case and consequences of different types and length of punishments. Given the complexity of the decision, it is natural that the personality and experience of the judge affect what sentence he will choose. It is right that criminal justice is built on the independence of judges and the principle of individualization of sentences.

It has become a widespread topic in the public discussion that more specified sentencing principles that help judges in sentencing decisions may limit the role of the personality of the judge and thus promote refinement of sentencing. In particular, it may lead to, among others establishing non-binding instructions on how to proceed in a sentencing decision. Such instructions may not only help judges in the sentencing decisions, but also provide a better understanding of the type of punishment and why was imposed by offenders and the general public.

We, the signatories of this petition, would like to invite members of the Committee

on Constitutional and Legal Affairs of the Chamber of Deputies, Parliament of the Czech Republic, to support our action.

Debriefing Letter

Last week you participated in a survey, in which we – on behalf of our client – informed you about shares of cases in which judges of one of the regional courts in the Czech Republic sentence offenders to community service for a crime of *failure to pay alimony*.

The information that was showed to you is truthful and describes a situation at one of the regional courts. However, the situation may not correspond to other regional courts. In the Czech Republic, there are several dozens of regional courts that may vary in differences among judges in the propensity to sentence offenders to different types of punishments for different crimes.

Should you be interested in sentencing decision in the Czech Republic, there is a webapage *jaktrestame.cz* devoted to it.

Bibliography

Acemoglu, D., Cheema, A., Khwaja, A. I., and Robinson, J. A. (2020). Trust in State and Nonstate Actors: Evidence from Dispute Resolution in Pakistan. *Journal of Political Economy*, 128(8):3090–3147.

Aizer, A. and Doyle, J. J. (2015). Juvenile Incarceration, Human Capital, and Future Crime: Evidence From Randomly Assigned Judges. *Quarterly Journal of Economics*, 130(2):759–804.

Alesina, A. and La Ferrara, E. (2002). Who Trusts Others? *Journal of Public Economics*, 85(2):207–234.

Anderson, J. M., Kling, J. R., and Stith, K. (1999). Measuring Interjudge Sentencing Disparity: Before and After the Federal Sentencing Guidelines. *The Journal of Law and Economics*, 42(S1):271–308.

Ashworth, A. (2009). Techniques for Reducing Sentence Disparity. In von Hirsch, A., Ashworth, A., and Roberts, J. V., editors, *Principled Sentencing: Readings on Theory and Policy*, pages 243–257. Hart Publishing, Oxford and Portland, OR, 3rd edition.

Athey, S. and Imbens, G. W. (2017). The Econometrics of Randomized Experiments. In *Handbook of Economic field Experiments*, volume 1, pages 73–140. Elsevier.

Barrie, C. (2020). Searching Racism After George Floyd. *Socius*, 6:2378023120971507.

Besley, T., Folke, O., Persson, T., and Rickne, J. (2017). Gender Quotas and the Crisis of the Mediocre Man: Theory and Evidence from Sweden. *American Economic Review*, 107(8):2204–2242.

Beyer, A., Cohen, D. A., Lys, T. Z., and Walther, B. R. (2010). The Financial Reporting Environment: Review of the Recent Literature. *Journal of Accounting and Economics*, 50(2-3):296–343.

Bjerk, D. (2005). Making the Crime Fit the Penalty: The Role of Prosecutorial Discretion under Mandatory Minimum Sentencing. *The Journal of Law and Economics*, 48(2):591–625.

Bjerk, D. (2017a). Mandatory Minimum Policy Reform and the Sentencing of Crack Cocaine Defendants: An Analysis of the Fair Sentencing Act. *Journal of Empirical Legal Studies*, 14(2):370–396.

Bjerk, D. (2017b). Mandatory Minimums and the Sentencing of Federal Drug Crimes. *The Journal of Legal Studies*, 46(1):93–128.

Blinder, A. S., Ehrmann, M., Fratzscher, M., De Haan, J., and Jansen, D.-J. (2008). Central Bank Communication and Monetary Policy: A Survey of Theory and Evidence. *Journal of Economic Literature*, 46(4):910–45.

Buisseret, P., Folke, O., Prato, C., and Rickne, J. (2019). Party Nomination Strategies in List Proportional Representation Systems.

Caillaud, B. and Tirole, J. (2008). Parties as Political Intermediares. *Quarterly Journal of Economics*, 117(4):1453–1489.

Carlsson, F., Demeke, E., Martinsson, P., and Tesemma, T. (2018). Measuring Trust in Institutions.

Chapman, B., Mirrlees-Black, C., and Brawn, C. (2002). *Improving public attitudes to the Criminal Justice System: The impact of information*. Home Office London.

Cheema, A., Hameed, Z., and Shapiro, J. N. (2017). Victimization, Citizen Engagement, and Policing in Lahore. *Institute of Development and Economic Alternatives, Policy Report, Lahore, Pakistan*.

Cirone, A., Cox, G. W., and Fiva, J. H. (2020). Seniority-based Nominations and Political Careers. *American Political Science Review*, pages 1–18.

Clarkson, C. and Morgan, R. (1995). *The Politics of Sentencing Reform*. Clarendon Press, Oxford.

Corda, A. (2016). Sentencing and Penal Policies in Italy, 1985 – 2015: The Tale of a Troubled Country. *Crime and Justice*, 45(1):107–173.

Council of Europe (1992). Recommendation No. R (92) 17 of the Committee of Ministers to Member States Concerning Consistency in Sentencing. Technical report, Council of Europe, Committee of Ministers.

Coutts, A. (2019). Good News and Bad News Are Still News: Experimental Evidence on Belief Updating. *Experimental Economics*, 22(2):369–395.

Crabtree, S. (2020). Most Americans Say Policing Needs 'Major Changes.'. *Gallup, July*, 22.

Dahl, G. B., Kostol, A. R., and Mogstad, M. (2014). Family Welfare Cultures. *Quarterly Journal of Economics*, pages 1–42.

Dal Bó, E., Dal Bó, P., and Snyder, J. (2009). Political Dynasties. *Review of Economic Studies*, 76(1):115–142.

Dal Bó, E. and Finan, F. (2018). Progress and Perspectives in the Study of Political Selection. *Annual Review of Economics*, 10(1):541–575.

Dal Bó, E., Finan, F., Folke, O., Persson, T., and Rickne, J. (2017). Who Becomes a Politician? *Quarterly Journal of Economics*, 132(4):1877–1914.

Dhami, M. K., Belton, I. K., Merrall, E., McGrath, A., and Bird, S. M. (2020). Criminal Sentencing by Preferred Numbers. *Journal of Empirical Legal Studies*, 17(1):139–163.

Di Tella, R. and Schargrodsky, E. (2013). Criminal Recidivism after Prison and Electronic Monitoring. *Journal of Political Economy*, 121(1):28–73.

Drápal, J. (2020). Ukládání trestů v případě jejich kumulace: Jak trestat pachatele, kteří spáchali další trestný čin předtím, než vykonali dříve uložené tresty. *Jurisprudence*, (2):1–17.

Drápal, J. and Dušek, L. (2021). Law or Policy? The Role of Informal Authority in Criminal Sentencing. *Unpublished manuscript*.

Eil, D. and Rao, J. M. (2011). The Good news-bad news Effect: Asymmetric Processing of Objective Information About Yourself. *American Economic Journal: Microeconomics*, 3(2):114–38.

Eurobarometer, S. (2018). Public Opinion in the European Union.

Ferraz, C. and Finan, F. (2009). Motivating Politicians: The Impacts of Monetary Incentives on Quality and Performance. *NBER Working Paper Series*, (No. w14906).

Fiva, J. H. and Røhr, H. L. (2018). Climbing the Ranks: Incumbency Effects in Party-list Systems. *European Economic Review*, 101:142–156.

Fleetwood, J. (2011). Five Kilos: Penalties and Practice in the International Cocaine Trade. *British Journal of Criminology*, 51(2):375–393.

Foulds, J. A. and Nutt, D. (2020). Principled Sentencing for Drug Supply Offences: Revised Methamphetamine Sentencing Guidelines in New Zealand. *Drug Science, Policy and Law*, 6:205032452094234.

Frankel, M. E. (1972). Lawlessness in Sentencing. *University of Cincinnati Law Review*, 41(1):1–54.

Galasso, V. and Nannicini, T. (2011). Competing on Good Politicians. *American Political Science Review*, 105(01):79–99.

Galasso, V. and Nannicini, T. (2015). So closed: Political Selection in Proportional Systems. *European Journal of Political Economy*, 40:260–273.

Galasso, V. and Nannicini, T. (2017). Political Selection Under Alternative Electoral Rules. *Public Choice*, 171(3-4):257–281.

Galil, K. and Soffer, G. (2011). Good news, Bad News and Rating Announcements: An Empirical Investigation. *Journal of Banking & Finance*, 35(11):3101–3119.

Grimmelikhuijsen, S. and Klijn, A. (2015). The Effects of Judicial Transparency on Public Trust: Evidence from a Field Experiment. *Public Administration*, 93(4):995–1011.

Grönlund, K. and Setälä, M. (2012). In Honest Officials We Trust: Institutional Confidence in Europe. *The American Review of Public Administration*, 42(5):523–542.

Hibbard, J. H., Stockard, J., and Tusler, M. (2005). Hospital Performance Reports: Impact on Quality, Market Share, and Reputation. *Health Affairs*, 24(4):1150–1160.

Hofer, P. J. (2019). Federal Sentencing after Booker. *Crime and Justice*, 48:137–186.

Jackson, J., Asif, M., Bradford, B., and Zakria Zakar, M. (2014). Corruption and Police Legitimacy in Lahore, Pakistan. *British Journal of Criminology*, 54(6):1067–1088.

Johnson, B. D. (2019). Trials and Tribulations: The Trial Tax and the Process of Punishment. *Crime and Justice*, 48:313–363.

Jurajda, Š. and Münich, D. (2015). Candidate Ballot Information and Election Outcomes: The Czech Case. *Post-Soviet Affairs*, 31(5):448–469.

Kaspar, J. (2020). Sentencing Guidelines vs. Free Judicial Discretion - Is German Sentencing Law in Need of Reform. In Ambos, K., editor, *Sentencing: Anglo-American and German Insights*, pages 337–352. Göttingen University Press, Göttingen.

Kert, R., Slot, B., Vernimmen, G., Lehner, A., and Schellekens, J. (2015). Study on Minimum Sanctions in the EU Member States. Technical report, European Commission.

Ketelaar, N. A., Faber, M. J., Flottorp, S., Rygh, L. H., Deane, K. H., and Eccles, M. P. (2011). Public Release of Performance Data in Changing the Behaviour of Healthcare Consumers, Professionals or Organisations. *Cochrane database of systematic reviews*, (11).

Kling, J. R. (2006). Incarceration Length, Employment, and Earnings. *American Economic Review*, 96(3):863–876.

Leader-Elliott, I. (2012). Sentencing by Weight: Proposed Changes to the Commonwealth Code, 9.1 Serious Drug Offences. *Criminal Law Journal*, 36:265–281.

Lepage, L.-P. (2020). Are criminals strategic? Offender Responses to Drug Sentencing Cutoffs. *Labour Economics*, 66:101906.

MacQueen, S. and Bradford, B. (2015). Enhancing Public Trust and Police Legitimacy During Road Traffic Encounters: Results from a Randomised Controlled Trial in Scotland. *Journal of experimental criminology*, 11(3):419–443.

Mattozzi, A. and Merlo, A. (2015). Mediocracy. *Journal of Public Economics*, 130:32–44.

Morris, S. and Shin, H. S. (2002). Social Value of Public Information. *American Economic Review*, 92(5):1521–1534.

Moutsiana, C., Garrett, N., Clarke, R. C., Lotto, R. B., Blakemore, S.-J., and Sharot, T. (2013). Human Development of the Ability to Learn from Bad news. *Proceedings of the National Academy of Sciences*, 110(41):16396–16401.

Murphy, K., Mazerolle, L., and Bennett, S. (2014). Promoting Trust in Police: Findings from a Randomised Experimental Field Trial of Procedural Justice Policing. *Policing and Society*, 24(4):405–424.

Palguta, J. (2015). Political Rent-Seeking in Public Procurement: Evidence from the Entry of Political Challengers at Electoral Thresholds. *CERGE-EI Working Paper Series*, (549).

Palguta, J. and Pertold, F. (2018). Do Higher Wages Produce Career Politicians? Evidence from Two Discontinuity Designs. *CERGE-EI Working Paper Series*, (630).

Pease, K. and Sampson, M. (1977). Doing Time and Marking Time. *How. J. Penology & Crime Prevention*, 16:59.

Sentencing Council (2011). Drug Offences Guideline: Professional Consultation. Technical report.

Shermer, L. O. and Johnson, B. D. (2010). Criminal Prosecutions: Examining Prosecutorial Discretion and Charge Reductions in U.S. Federal District Courts. *Justice Quarterly*, 27(3):394–430.

Skugarevskiy, D. (2017). *Essays in Law and Economics of Enforcement*. Dissertation thesis, Graduate institute of international and development studies in Geneva.

Smith, P. C., Mossialos, E., Papanicolas, I., and Leatherman, S. (2009). *Performance measurement for health system improvement: experiences, challenges and prospects*. Cambridge University Press.

Sporer, S. L. and Goodman-Delahunty, J. (2009a). Disparities in Sentencing. *Social Psychology of Punishment of Crime*, pages 379–401.

Sporer, S. L. and Goodman-Delahunty, J. (2009b). Disparities in Sentencing Decisions. *Social Psychology of Punishment of Crime*, pages 379–401.

Sutela, M. (2020). Rangaistusten jakautuminen rangaistusasteikkojen rajoissa oikeuskäytännössä. *Defensor Legis*, (2):215–229.

Titl, V. and Geys, B. (2019). Political Donations and the Allocation of Public Procurement Contracts. *European Economic Review*, 111:443–458.

Tonry, M. (2019). Fifty Years of American Sentencing Reform: Nine Lessons. *Crime and Justice*, 48:1–34.

Travova, E. (2019). Under Pressure? Performance Evaluation of Police Officers as an Incentive to Cheat: Evidence from Drug Crimes in Russia. *CERGE-EI Working Paper Series*, (637).

Tuttle, C. (2019). Racial Disparities in Federal Sentencing: Evidence from Drug Mandatory Minimums.

Ulmer, J. T. (2012). Recent Developments and New Directions in Sentencing Research. *Justice Quarterly*, 29(1):1–40.

Ulmer, J. T. and Bradley, M. S., editors (2019). *Handbook on Punishment Decisions: Locations of Disparity*. Routledge.

United States Sentencing Commission (2011). 2011 Report to the Congress: Mandatory Minimum Penalties in the Federal Criminal Justice System. Technical report.

Vaughn, P., Peyton, K., and Huber, G. A. (2021). Public Support for Different Movements to Reshape American Policing is Contingent on the Perceived Implications for Crime and Public Safety.

von Hirsch, A. (2017). *Deserved Criminal Sentences*. Hart Publishing, London.

Zeman, P. (2020). Vězeňství pohledem Nejvyššího státního zastupitelství. In Gřivna, T. and Šimánová, H., editors, *Ukládání trestů a jejich výkon*, pages 55–64. Aleš Čeněk, Plzen.