Sentencing Disparity and Attitude Towards Judicial System: Evidence from a Survey Experiment*

Michal Šoltés†

Preliminary and incomplete draft

The most recent version

Thursday 3rd September, 2020

Abstract

This paper studies whether information about sentencing disparity undermines public trust in the judicial system and affects (intended) courts related behavior in an environment of a relatively high level of trust. In particular, I conduct a survey experiment in which I provide the general public with varying information about sentencing disparity among judges in the Czech Republic. I show that information about sentencing disparity increases willingness to read and sign a petition that invites politicians to particularize sentencing principles by 4.3 percentage points (7 %). The effect is driven by: (i) female respondents with children who are more likely to be concerned about the sentencing disparity in the presented crime of *failure to pay alimony*; and (ii) individuals who have a priori low level of approval for the judicial system. Additionally, the information shifts policy preferences; fairness of the judicial system is viewed as higher-priority policy among respondents informed about the sentencing disparity. On the contrary, I find no evidence that the information about sentencing disparity affects declared institutional trust, willingness to apply to the court, nor demand for alternative dispute resolutions.

JEL Codes: C90, D02, K40.
Keywords:

*I thank Jakub Drápal. This study was supported by Charles University, GAUK projects No. 218218 and by a grant from the CERGE-EI Foundation's Teaching Fellows program. This project also received funding from the European Research Council under the European Union's Horizon 2020 research and innovation programme (grant agreement No 678081). All opinions expressed are those of the author(s) and have not been endorsed by the CERGE-EI Foundation. Registry: AEARCTR-0005374

†Email address: msoltes@cerge-ei.cz; CERGE-EI, a joint workplace of Charles University and the Economics Institute of the Czech Academy of Sciences, Politických vězňů 7, 111 21 Prague, Czech Republic.; Charles University, Faculty of Law, nám. Curieových 901/7, 116 40 Prague, Czech Republic.

1 Introduction

Discretion in public authorities decision-making often leads to significant disparities in their decisions. Sentencing disparity among judges — who systematically differ in their sentencing patterns (e.g. propensity to incarcerate) — is one of the most distinctive and worrying examples documented and discussed in academic literature and media.[1] The fact that, in spite of formal rules of standard democracies (e.g., equality before the law; clear, stable, predictable application of the law), a personality of the judge assigned to a case plays a key role in the sentencing decisions, may undermine perceived competence of the judicial system to act right and fair, and thus weaken the public trust.

Justifiable trust in institutions often comes with economic benefits through a more effective provision of goods and services. In comparison to private consumers, (trusted) institutions provide public services such as law and contract enforcement with a lower cost. It thus has become a concern of policy makers to (re)build and maintain institutional trust, most notably since the 2008 financial crises and subsequent waves of populism that have contributed to an erosion of public trust (OECD, 2017). Importantly, recent years have also witnessed an increase of data and information availability, including information about performance of public institutions that may affect the perceived performance of institutions and thus the level of trust. Therefore, understanding of whether and through which channels information about sentencing disparity (or any other weak performance of public institutions) may cause reaction of the general public has intelligible policy implications. Recently, Acemoglu et al. (2020) documented that truthful information about performance of a public institution alone has the power to affect citizens attitude towards the public institution with economic and social consequences.

In this paper, I study whether and how information about sentencing disparity affects attitude towards the judicial system. In particular, I conduct a survey experiment with 2410 respondents who were provided with varying yet not deceptive information about sentencing disparity among judges at regional courts in the Czech republic. In a between subjects design I measure the effect of information on: (i) declared institutional trust; (ii) courts related behavior such as willingness to apply to courts and demand for alternative dispute resolutions; and (iii) policy preferences regarding the judicial system. The information provided consists of shares of cases in which convicted offenders were sentenced to community services as opposed to different types of punishments (e.g. incarceration) for one of the most frequent crimes - *failure to pay alimony*. The source of the variation

[1] A well-known example is an analysis published by Reuters in 2017, that "found broad variations among judges and courts in how frequently immigrants are deported. Houston, Charlotte and Atlanta have been labeled "asylum free zones" by immigrant advocates who say those courts disproportionately deny claims for protection. Within specific courts, too, rulings vary widely from judge to judge." `https://www.reuters.com/investigates/special-report/usa-immigration-asylum/`

comes from the different levels of sentencing disparities (i.e. shares of cases sentenced to community service) among judges within a given court across different courts; while at one court, the propensity to sentence to community service among judges is similar, at a different court the sentencing pattern shows significant differences. The treatment group was informed about a situation at a court with significant sentencing disparity, whereas the active control group was informed about a court at which all judges tend to choose community service with the same propensities.

I show that information about sentencing disparity shapes policy preferences. First, respondents were informed about a petition that invites politicians to suggest particularizing sentencing principles which would assist judges in their sentencing decisions and asked whether they want to read and sign such a petition. The information about sentencing disparity increases the respondents' willingness to read and sign the petition by 4.3 percentage points (7 %). Second, the respondents were asked to rank the following policy issues: (i) fairness of the judicial system; (ii) sufficient highway infrastructure; (iii) high-quality of teachers in the education system; (iv) safety situation in the Czech Republic, according to the priority they would address them should they become the prime minister. Fairness of the judicial system was 3.6 percentage points (11.5 %) more likely to be considered as the top priority by the treatment group than by the control group. Additionally, I find evidence that the effects are driven mainly by two groups of respondents: (i) female respondents with at least one child, i.e. those who are arguably most sensitive to the criminal cases of *failure to pay alimony*; and (ii) those with an a priori low level of approval of the judicial system. The former suggests that the response to the information about sentencing disparity is likely crime- and/or situation-specific.

I do *not* find any treatment effects on other outcomes studied. In particular, the information does *not* change the declared level of institutional trust neither in judicial system nor in other relevant institutions (e.g. the police). Similarly, the information has *no* effect on the respondents' willingness to apply to the court in a situation in which they were asked by an NGO to give their advice to a single mother who questions the purport of applying to the court. Information about sentencing disparity does *not* affect respondents' demand for alternative dispute resolutions (ADR) measured by respondents' interest in an offer to receive a free booklet with information about ADR. Due to relatively low standard errors, I interpret the estimates as evidence of null results rather than imprecise estimates of potentially significant effects.

TODO: discussion paragraph

This project shares several features with a recent paper by Acemoglu et al. (2020) in which the authors show that truthful information about reduced delays in state courts increases reported likelihood of using formal courts as opposed to non-state institutions

(Panchayats) in rural Pakistan. Their study differs diametrically in the context form mine: (i) the project by Acemoglu et al. (2020) was conducted in a rural area in Pakistan where households access the court system frequently[2] 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 the performance of courts is less widespread and respondents are less experienced in judicial and courts' practice. In fact, only 10 % of respondents reported that they have had sizeable experience with the judicial system (first-hand and/or through people they know well, e.g. family)[3]; (ii) the judicial systems in Pakistan and in the Czech Republic enjoy different level of public trust. According to the Eurobarometer (2018), in the Czech Republic 43 % of respondents tend to trust in the judicial system which is by 8 percentage points fewer than the average of the EU28, yet still comparable with most of the developed countries. On the contrary, in Pakistan the state institutions suffer from a lack of trust (Jackson et al., 2014; Cheema et al., 2017).[4]

More broadly, I contribute to three streams of literature. First, this paper extends a body of literature devoted to (in)consistency in sentencing and sentencing disparities by adding evidence on the consequences of existence of sentencing disparity on the attitude of the general public towards the judicial system. 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, 2009).[5] Many scholars have leveraged the different practice (leniency) of judges as a source of quazi-exogenous variation.[6] For example, Kling (2006) uses differences in sentencing propensities among judges to estimate the causal effect of

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

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

[4]"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)

[5]Several different mechanisms have been argued to contribute to the disparities, e.g. variations in sentencing philosophies held by different judges; variability that arises from attitudes toward specific types of crimes and interactions with a judge's personal experience (e.g. a judge with teenage daughters who must sentence a repeat offender regarding child sexual abuse); the appearance of the offender, such as attractiveness, including race, ethnicity, and gender; other factors unrelated to the merit of the cases such as having a trial on birthdays may contribute to inconsistent sentencing (Sporer and Goodman-Delahunty, 2009; Mustard, 2001; Steffensmeier and Demuth, 2006; Chen and Philippe, 2017)

[6]Similarly, others use different patterns in practice (leniency) of authorized decision-makers such as police officers about speeding tickets (Goncalves and Mello, 2017), IMF Mission Chief (Beaudry and Willems, 2018), and patent examiners (Gaulé, 2018; Sampat and Williams, 2019).

incarceration on labor market outcomes. Di Tella and Schargrodsky (2013) study recidivism using ideological differences across judges that drive the likelihood of being sent to electronic monitoring instead of prison differs substantially across judges. Furthermore, applying a similar identification strategy, Dahl et al. (2014) investigate the heredity of welfare program participation, Aizer and Doyle (2015) study the consequences of incarceration of juveniles on completion of high-school and future criminal behavior, Dobbie et al. (2018) study the effect of pretrial detention on the probability of conviction, and Bhuller et al. (2020) as well as Mueller-Smith (2015) estimate a causal effect of incarceration on recidivism and labor market outcomes. My project differs from most of the projects, as it uses the variation as a source of information in a survey experiment, rather than a variation for quazi-experimental approaches. To leverage the variation in the experimental design, I apply an approach similar to: (i) Bérgolo et al. (2019), who introduce exogenous, but non-deceptive variation in the information about audit probabilities and penalty rates in the audit-statistics letter; and (ii) Bursztyn et al. (2019) who exogenously manipulate information whether Donald Trump or Hillary Clinton won the Presidential election in "Pittsburgh", by manipulating the exact area, as Donald Trump won in "Pittsburgh's metropolitan area" while Hillary Clinton in "Pittsburgh's county."

Second, this project also belongs to the stream of literature using information provision experiments to study a role of beliefs about important social issues on behavior of individuals. Haaland et al. (2020) review methodological questions, discuss common practices, and provide an extensive literature review of the information provision experiments in several fields of economics. To highlight some of the studies, Kuziemko et al. (2015) inform respondents about inequality and show that the information treatment has large effects on views about inequality but only slightly affects policy preferences, Alesina et al. (2018) and Grigorieff et al. (2016) study the impact of information treatment on the attitude to immigrants. Haaland and Roth (2018) study beliefs about racial discrimination and document that the information provided leads to belief updating. Haaland and Roth (2017) find that labor market concerns drive support for immigration. Cappelen et al. (2019) study beliefs about behavioral responses to taxation. Cruz et al. (2018a,b) study the effect of information treatment on voting behavior. I see my contribution to this literature as twofold. First, while most of the previous project inform subjects about the first moment of a distribution, e.g. share of immigrants, the core of the treatment is to inform about sentencing disparity, i.e. the second moment of a distribution of sentencing decisions, which is connected with a few challenges such as unintentional variations in the perception of the first moment and complicated elicitation and quantification of prior beliefs.[7] Second, I extend the literature by studying a novel research question using

[7]One cannot simply ask what is the perceived standard deviation in propensity to sentenced to

the information provision experiment.

Third, this paper also contributes to literature studying institutional trust, specifically, the role of perceived quality of institutions on the declared level of trust. Unlike the interpersonal trust which is usually elicited using economic 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. In fact, Carlsson et al. (2018) and Alesina and La Ferrara (2002) find a low correlation between generalized trust and trust in different institutions (government, police, judicial system). Many scholars studying institutional trust rely on declared institutional trust data and provide results limited to correlational evidence (e.g., Grönlund and Setälä, 2012). Evidence on the causal effect of interventions on institutional trust is generally rare. Two projects that aim to estimate the causal effect on institutional trust studied the effect of procedural justice protocol on trust in the police. Murphy et al. (2014) argue that if police follow an experimental protocol — that focuses on voice, neutrality, trustworthiness, and respect — during a control, drivers in Australia reported higher trust in police; however, using a similar experimental design in Scotland, MacQueen and Bradford (2015) fail to replicate the effect of an increase in trust. Similar question of whether transparency of judicial system affects institutional trust was studied by Grimmelikhuijsen and Klijn (2015). Specifically, they conduct a field experiment in the Netherlands in which they invited respondents to watch a TV series about a District Court in the Netherlands that allowed the public to watch daily work of judges on real cases. The authors report that watching the TV series increases the declared level of trust in judges. The treatment in (Grimmelikhuijsen and Klijn, 2015), however, conveys different information than the one in this paper. 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 performance of public institution on declared institutional trust and, importantly, other measures of indented behaviors that are likely related to the trust.

The rest of the paper is organized as follows. The following section introduces the design of the experiment and the outcomes studied. Next, I discuss the results, both the average treatment effect and the heterogeneous treatment effects. Finally, before I conclude, I add thought on interpretation, generalizability, and implications of the results.

community services among judges.

2 Design of Survey Experiment

To conduct the survey experiment I partner 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. 2410 respondents completed the survey. Additional to the data collected in the experiment, I have basic demographic characteristics of respondents collected in previous surveys. With the exception of the final task, the survey experiment was run on a platform of Behavio which is familiar to the respondents.

The experiment consists of four stages. First, upon starting the survey respondents were asked three questions regarding their prior attitude to, and previous experience with, the judicial system. Second, respondents were randomized into the treatment and the active control groups and were presented with the treatment and control slides, respectively. Third, after the information treatment respondents complete five tasks (questions). Finally, for those interested, the final task consists of reading and signing a petition that was posted on a different website and thus the respondents had to leave the Behavio platform. The order of tasks in the first and the third stages were randomized at the individual level. The full script is available in the online appendix.

2.1 Attitude

The aim of the first stage was to understand the respondents' initial attitude towards the judicial system. To do so, respondents were asked to what extent on a four-level scale they agree with two statements: (i) "Depending on a 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 closed to them are.

2.2 Treatment

I provide 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 control group about a court with negligible sentencing disparity among judges. The source of the variation comes from the fact that courts differ in their level of disparities among judges; in some regional courts judges vary in their sentencing patters significantly, while in others all judges exhibit indistinguishable sentencing patterns. The information used relies on variation within a given court, rather than between courts or among all judges across all courts, as some of the regional disparities in sentencing are justifiable and do not represent the

intended variation.[8]

Data about sentencing decisions are complicated and multidimensional (different types of crime, different types and extent of punishment, and combination of more types of punishment) which makes it complicated to convey an understandable message. Additionally, requirements to base the information treatment on sound empirical data restrict the possible content of the information to only a few of frequent and general types of crimes for which: (i) there are enough observations per judge that the differences in sentencing is statistically significant; and (ii) cases are assigned to judges at random. I focus on one of the most common crimes in the Czech republic – *failure to pay alimony*[9] – and present shares of cases in which a judge sentences convicted person to community services as the main punishment. In the Czech Republic, *failure to pay alimony* is a crime for which the convicted person can be given one or more of the following punishments: suspended sentence (69 %); community service (16 %); incarceration (13 %); house arrest (0 %)[10]. *Failure to pay alimony* is a suitable type of crime, as it satisfies the aforementioned criteria: (i) I can provide aggregate statistics based on at least 80 cases per judge (over a span of three years of 2016-2018); (ii) compared to other types crimes, 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 can be assigned to judges at random and importantly, according to the courts' schedule, they indeed are.

The treatment and the control slides present the information in a style news would do. The core of the slide is a simple self-explanatory bar graph accompanied by a few additional pieces of information that provides 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 the judge C (29 %) sentenced to community service almost a third of the convicted offenders, whereas for some it is less than 10 % and instead they apply difference types of punishments. Finally, the slide highlights that cases are assigned at random, and that being assigned to the judge C implies up to 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

[8]For example, losing a driving license/a driving disqualification in a city with functional public transportation is a less strict punishment in terms of economic and social consequences than losing a driving license in regions at the foothills of mountains with limited public transportation.

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

[10]Frequency of different types of punishment is calculated using data from 2016 - 2018.

further explains that regardless the judge assigned, a convicted offender has very similar probability of being sentenced to community service.[11][12]

2.3 Experimental Outcomes

The outcomes collected are classified in 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 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 led by a connection to the judicial system. The closest institution to the judicial system is the police, as police officers often cooperate on criminal cases. Next institution is the government which is responsible for functional judicial system[13]; and finally the public broadcasting, that can be viewed responsible for a lack of information about sentencing disparity.

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

To understand whether providing information about sentencing disparity reduces willingness to apply to the court, I cooperate with a NGO (*vasevyzivne.cz*), that assists single-parents to sue out for alimony. In the experiment, I briefly explain situation of a client of the NGO and that she questions whether to apply to the court or nor. Applying is potentially beneficial, but it also may lead to high cost, both in terms of money and time and no benefits. Then, I asked the respondents whether they would recommend her to apply to the court or not and informed them that their advice may be used by the NGO

[11] The screens of the control and the treatment slides can be found in the online appendix.

[12] Once respondents finished the experiment, the Behavio sent them debriefing letter that informs them about 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.

[13] For example, the government is responsible for nomination of judges, so the President can appoint them.

as a material in similar cases. Presumably, the belief that their responses will potentially serve as guideline for people in actual problems increases the cost of ill-concerned answer.

If information about sentencing disparity would 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 fact, Acemoglu et al. (2020) document substitutability between formal and informal courts motivated by perception of a poor performance of the formal courts. As the next task, I explained that in some cases, it is possible to rely on alternative dispute resolutions instead of the judicial system and then I offered a free booklet with basic information about the alternative dispute resolutions. They first provided an indicate answer of their interest and if it was affirmative they were asked to provide their email addresses to have the booklet sent. The two-steps procedure was done for two reasons. First, it evaded legal concerns about using their email addressed for different purposes than inviting them to conduct the survey and second, it imposed cost on the action.

Policy Preferences Previous literature and and real-life experience document that new information about how public institutions work causes changes in policy preferences and may evoke public reaction (e.g. petitions and political protests). To measure the effect I collect two outcomes regarding the policy preferences.

First, I asked respondents to imagine that they are the prime minister of the Czech Republic and then gave them a list of four policy issues that were supposed to be ranked according to the perceived priority. The acutest issue was supposed to be ranked as the top priority, the second most acutest as the second priority and so on. The four policy issues were: (i) fairness of judicial system; (ii) sufficient highway infrastructure; (iii) high-quality teachers in educational systems; and (iv) security situation in the Czech Republic.

Second, I elicited respondents' willingness to become actively involved in policy debate. In particular, I utilize academic discussion about whether judges would benefit from having more particularized sentencing principles at their disposal and had prepared a petition that invites politicians (members of the Committee on Constitutional and Legal Affairs, Chamber of Deputies, Parliament of the Czech Republic) to suggest particularizing sentencing principles. Respondents in the experiment were briefly informed about the content of the petition and asked if they would be interested in reading and signing the petition. If interested, the respondents were refer to a google forms website with the full text of the petition to read. The petition says that sentencing decisions are extremely important in one's life with far-reaching consequences and it may be beneficial to introduce a manual that helps to unify sentencing. Importantly, the petition was explicit that

the manual would be designed to help judges in their sentencing decisions, but would not undermine their independence. If interested in signing the petition, the respondents could have left an email address to have the signature sheet sent.[14] 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 do not observe responses at the individual level. However, since individuals in treated and control groups were referred to different form of the identical petition, I observe the number of email addressed left for the control and the treatment group separately.

3 Results

In this section, I present results. I start with outcomes collected prior to randomization, then I move to experimental evidence and estimate the average treatment effects on three different groups of outcomes and finish with heterogeneity of the treatment effect.

Data collected through online survey experiments are prone to suffer from low attention of subjects and careless responses. To overcome the issue, I restrict the dataset by disposing of observations from respondents who did not pay enough attention to the treatment and control slides, measured by the time spent on the slides. The treatment (control) slide consists of four (two) pieces of information and a graph, it thus seems reasonable to expect that one needs at least about 20 (15) seconds to understand the meaning of the slide. 20 (15) seconds represent 16^{th} (17^{th}) percentile of time spent on the corresponding slide. To obtain rather conservative estimates, for most analyses I dispose *only* of 10 % of observations with the least time spent on the slides. The 10^{th} percentile corresponds to 17 (13) seconds spent on the treatment (control) slide. Additionally, for several respondents it took extremely long to move on and I also suspect they did not pay attention to the task and responded carelessly. As a result, I do not include observations of respondents who spent more than the 98^{th} percentile respondent (approx. 3.3 of median time of treatment and control groups, respectively). I refer to the sample as the *baseline* sample.

3.1 Prior Attitude Towards Judicial System

A majority of respondents (91.4 %) agree that sentencing decisions are sensitive to a personality of a judge and that depending on a judge assigned, similar cases can be sentenced differently. The measure of general approval of the judicial system is less unequivocal, as 52.5 % agree that the judicial system works well, while 47.6 % are less

[14]The respondents were informed that the petition will be sent once there are at least 1000 signatures.

optimistic and disagree. The immediate implication is that many respondents (46.7 %) tend to approve of the judicial system, despite the perceived sentencing disparity.[15]

Prior beliefs regarding the judicial system, both about the level of approval and perception of sentencing disparity, vary by reported experience with the judicial system. Figure 1 shows a frequency of different combinations of prior approval (x-axis) and perception of sentencing heterogeneity (y-axis) by groups that report different experience with the judicial system. Among 10 % of respondents who reported sizeable experience with the judicial system (top left subfigure), most of them (60 %) strongly agree that depending on judge assigned, similar cases can be sentenced differently and they do not consider the judicial system working well. The remaining three groups of respondents with lower level of experience with the judicial system show different pattern. The majority consider the judicial system rather working and do not view the sentencing disparity as that extreme, yet they still admit it may exist.

Formally, Wilcoxon rank-sum test suggests that sizeable prior experience with the judicial system is associated with different attitude towards the judicial system. Testing the same level of approval of the judicial system between the group of the most experienced respondents and the remaining three groups pooled together, strongly suggests to reject the equality (p-value = 0.000). Similarly, the same comparison for perception of sentencing heterogeneity suggests rejection of the equality (p-value = 0.000). While the relationship between the level experience and the attitude is only correlational, it is an indication that the level of knowledge and understanding of the judicial system is related to the attitude towards the judicial system.

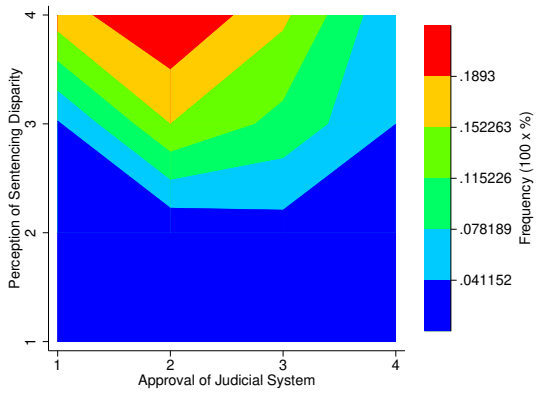
3.2 Experimental Evidence

3.2.1 Randomization

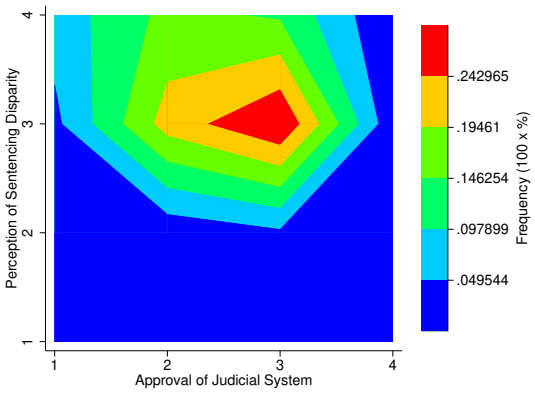
Respondents were randomized into two groups: (i) the *control* group, which was shown a control slide with a court with homogenous sentencing patterns among judges; and (ii) the *treatment* group, which was shown a slide with a court with a significant sentencing disparities among judges. The groups are balanced on both observed demographic characteristics and their prior attitudes towards the judicial system. The perception of information about the *failure to pay alimony*, i.e. the treatment information can vary with respondents' marital status and the fact whether they have children. 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. Additional 17 % (15 %) reported they are currently divorced and only 4% (3%) are widowed. The reported

[15] Figure 9 in the appendix shows the aggregate levels of responses.

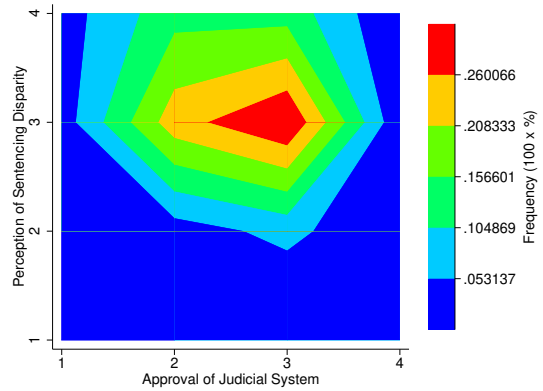
Figure 1: Attitude Towards Judicial System by Experience



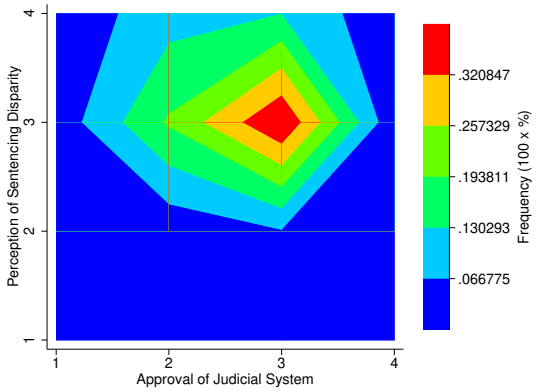
(a) Sizable Experience With the JS



(b) Not Sizeable Experience with the JS



(c) Superficial Experience with the JS



(d) None 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.

marital status reflects their status at the time of the survey experiment but not their history; for example, respondents reported as married could have experienced a divorce. Slightly more than 70 % of respondents have at least one child and the average number of children is 1.44 in the control group and 1.49 in the treatment group. In both groups, there are fewer male (46.6 % in the treatment group and 48.5% in the control group) than female respondents. For more details, see table 6 in the appendix.

3.2.2 Institutional Trust

From the four institutions studies, the police enjoys the highest level of trust (55 % of respondents report one of the two highest level of trust), followed by the judicial system (41 %). Respondents reported relatively low level of trust in the national government (20 %). Similarly, the trust in the public broadcasting (24 %) is also low. Figure 2 captures a frequency of different levels of institutional trust declared for all four institutions studied and shows preliminary evidence that the level of institutional trust does not depend on treatment status.

The levels of trust are consistent with other surveys of reported institutional trust. For example, institutional trust in the police, the national government, and the judicial system can be compared to the results from Eurobarometer (2018), in which respondents were given a list of institutions and asked whether they trust these institutions. The most trusted institution (of these three) is the police, as 63 % of respondents report that they tend to trust the police; followed by the judicial system (43 %); and the least trusted is the national government (28%). Overall, my results, while less optimistic, tend to resemble those from Eurobarometer (2018).[16].

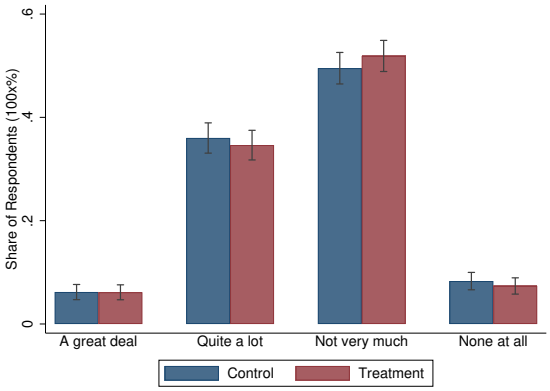
To test the effect of information about sentencing disparity on institutional trust, I conduct four exercises that all suggest that the treatment does not have an effect. I start by collapsing the first two possible answers: a great deal of trust; and quite of a lot of trust into one category denoted as $High\,Trust$. I use this variable to: (i) estimate univariate OLS regressions (1) for each institution

$$High\,Trust = \alpha + \beta\ Treatment + \varepsilon \tag{1}$$

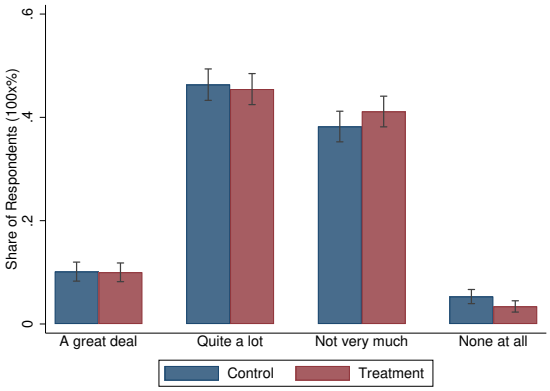
(ii) estimate OLS regressions with a set of dummies capturing individual prior attitude towards the judicial system (experience, randomness, and approval) and basic demographic characteristics (income, age, level of education, gender, and the number of kids); (iii) simulate the exact p-value for the sharp null hypothesis derived from the potential

[16]Eurobarometer (2018) did not ask about the institutional trust in the public broadcasting

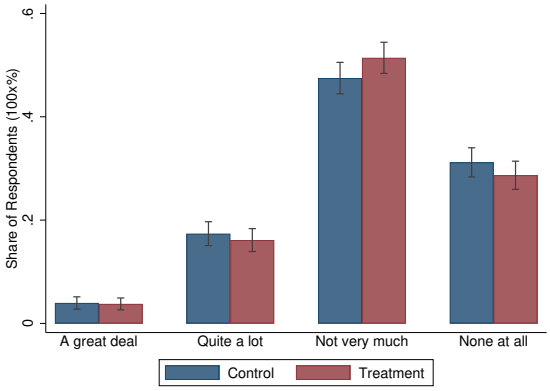
Figure 2: Levels of Institutional Trust



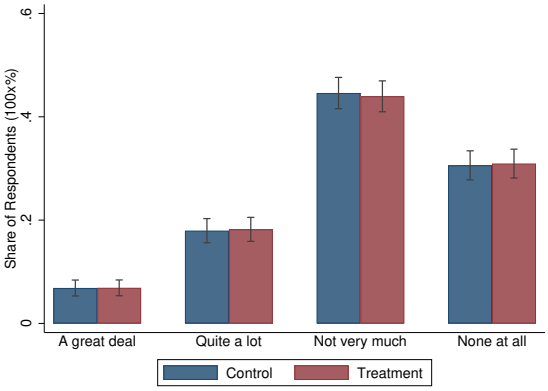
(a) Trust in Judicial System



(b) Trust in Police



(c) Trust in Government



(d) Trust in Public Broadcasting

Notes: Declared level of institutional trust by treatment status in (a) the judicial system; (b) in police; (c) government; and the public broadcasting. 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.

outcome framework (Athey and Imbens, 2017) and then test

$$High\,Trust_i(0) = High\,Trust_i(1) \quad \forall i = 1, \ldots, N. \tag{2}$$

Finally, since the levels of institutional trust represent 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 limited attention of respondents, I run the OLS regressions for two different samples: (i) the *baseline* sample, as described before; and (ii) a sample that drops observations for 25% of respondents who spent the least time on the institutional trust task. I refer to the latter as the *restricted* sample. Together with the observations dropped due to the limited attention on the treatment and control slides, the *restricted* sample counts for 67 % (1626 observations from 2410) of the most attentive respondents. If there is no effect in the *restricted* sample is thus credible evidence that the lack of effect is not caused by the limited attention of respondents.

Table 1 summarizes the results. Panel A and B show results from the OLS regressions and strongly suggest that the average treatment effect is economically and statistically insignificant. For all four institutions the point estimates of the treatment effect are, in absolute value, safely less than 2 percentage points. Proving information about sentencing disparities among judges thus did not affect the declared institutional trust. Results from Panel B further suggest that it is unlikely driven by respondents' inattention. Panel C uses the *baseline* sample and shows the exact p-value derived from 20,000 simulations and p-value of Wilcoxon rank-sum test. Panel C confirms the OLS results and provides additional evidence that the treatment effect is statistically insignificant.

3.2.3 Reliance on Judicial System

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 – a possible substitute to the judicial system – $ADR\ Interest$ and $ADR\ Mail$. Respondents exhibited a strong will to apply to the judicial system. In particular, almost 95 % of respondents recommended a single mother to apply to the court in a situation in which she hesitates. The effect of the treatment is negligible. At the same time, if offered, a majority (78 %) would be interested in receiving a booklet about alternative dispute resolutions. However, when the respondents were asked to provide their email addresses, so the booklet can be sent, only 24 % of all respondents and 31 % of those who declared their interest did so. The treatment information did not lead to any differences. See figure 3.

To test the average treatment effect formally, I conduct three exercises that all suggest

Table 1: Treatment Effect on Institutional Trust

Panel A: Baseline Sample

	Judicial System		Police		Government		Broadcasting	
Treatment	-0.014	-0.019	-0.010	-0.012	-0.014	-0.014	0.003	0.003
	(0.022)	(0.018)	(0.022)	(0.020)	(0.018)	(0.017)	(0.019)	(0.018)
Prior Attitude	No	Yes	No	Yes	No	Yes	No	Yes
Demo. Char.	No	Yes	No	Yes	No	Yes	No	Yes
N	2 096	2 093	2 096	2 093	2 096	2 093	2 096	2 093

Panel B: Restricted Sample

	Judicial System		Police		Government		Broadcasting	
Treatment	0.010	-0.006	0.010	0.002	-0.007	-0.008	0.013	0.009
	(0.024)	(0.020)	(0.025)	(0.023)	(0.021)	(0.020)	(0.021)	(0.021)
Prior Attitude	No	Yes	No	Yes	No	Yes	No	Yes
Demo. Char.	No	Yes	No	Yes	No	Yes	No	Yes
N	1 626	1 623	1 626	1 623	1 626	1 623	1 626	1 623

Standard errors in parentheses.

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

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

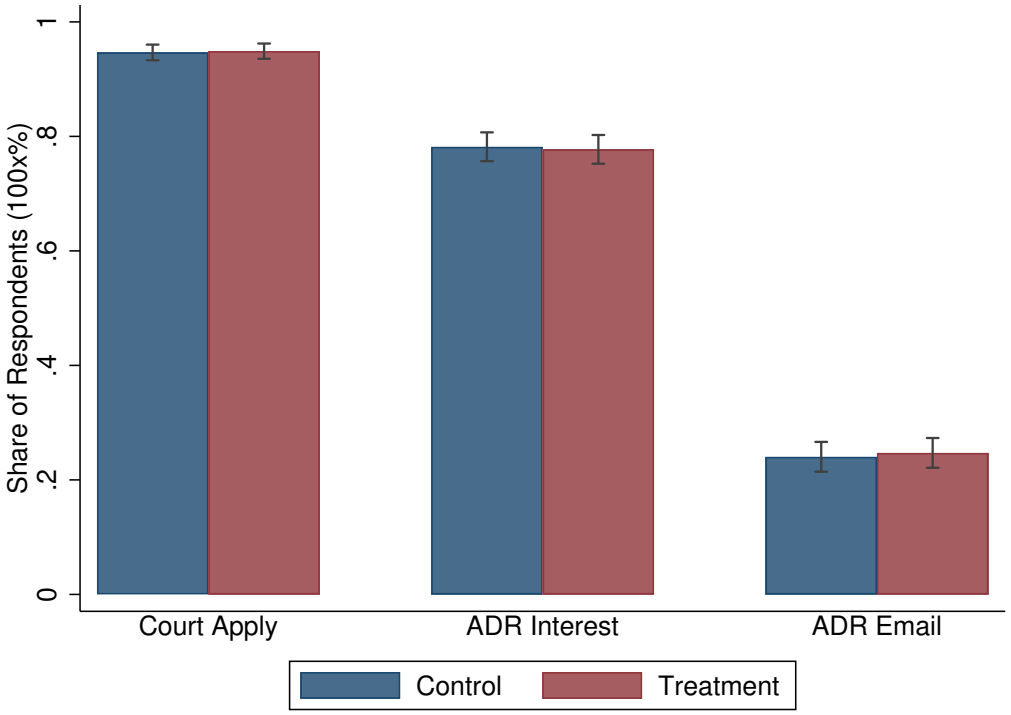
	Judicial System	Police	Government	Broadcasting
Exact p-value	0.505	0.648	0.418	0.879
Wilcoxon test	0.762	0.971	0.976	0.667

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 a sample without 25 % of respondents who spent the least time on the task.

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

Figure 3: Reliance on Judicial System



Notes: Shares of affirmative responds by the treatment status. *CourtApply* captures whether respondents recommend to apply to the court. *ADR Interest* measures shares of respondents who indicate their interest in receiving information about alternative dispute resolution (ADR). *ADR Mail* shows share of respondents who provide their email addresses to have the information about ADR sent.

that the providing information about sentencing disparity does not affect respondents' reliance on judicial system. For each of the three outcomes, I: (i) run univariate OLS regression; (ii) extend the OLS regression by controlling for prior attitude and demographic variables; (iii) simulate the exact p-value. I conduct all exercises on two samples: (i) the *baseline* sample; and (ii) on *restricted* samples that drop 25 % of respondents who spent the least time on the corresponding task. Note that since willingness to apply to the court and choice over alternative dispute resolutions are two different tasks, also the *restricted* samples differ.

Table 2 shows the results. Using the *baseline* sample, Panel A shows estimates of the treatment effect for each of the outcomes. The difference between the control and the treatment groups are statistically and economically insignificant. Panel B confirms the results on the *restricted* samples of respondents and provides thus additional evidence that the null effect is robust to inattention of respondents. Finally, Panel C presents exact p-values and suggests that none of the null hypotheses of no treatment effect cannot be rejected.

In terms of the magnitude of the point estimates, the effects are negligible and even smaller that the effect on institutional trust. With the exception of the $ADR\ Mail$ estimated on the restricted sample, all point estimates of the average treatment effects are bounded between -1 and 1 percentage points with standard deviations of 1 or 2 percentage points. The 95% confidence interval is thus bounded between -5 percentage points and 5 percentage points. I view these results as credible evidence of absence of the treatment effects rather than imprecise estimates of economically significant treatment effect.

3.2.4 Policy preferences

Information about sentencing disparity evokes response in both measures of policy preferences. First, respondents in the treatment group view fairness of judicial system as more important policy issue than respondents in the control group. Second, treated respondents are more likely to be interested in reading and signing a petition that invites politicians to particularized sentencing rules. At the level of individual respondents, I collect respondent's interest in reading and signing a petition and the rank of prioritized policy issues. Furthermore, at the group level – the control and the treatment groups – I collect the overall number of email addresses filled in by respondents who wish to receive the signature sheet to sign the petition.

Fairness of the judicial system is a relevant policy issue. Exactly one third (33.3 %) of respondents ranked fairness of the judicial system as the top priority and additional 50 % as the second most important priority. While the perception of fairness of the judicial

Table 2: Treatment Effect on Reliance on Judicial System

Panel A: Baseline Sample

	Court Apply		ADR Interest		ADR Mail	
Treatment	0.002	0.002	-0.004	-0.002	0.007	0.009
	(0.010)	(0.010)	(0.018)	(0.018)	(0.019)	(0.019)
Prior Attitude	No	Yes	No	Yes	No	Yes
Demo. Char.	No	Yes	No	Yes	No	Yes
N	2 086	2 083	2 096	2 093	2 096	2 093

Panel B: Restricted Sample

	Court Apply		ADR Interest		ADR Mail	
Treatment	-0.003	-0.004	-0.011	-0.008	0.003	0.007
	(0.011)	(0.011)	(0.021)	(0.020)	(0.021)	(0.021)
Prior Attitude	No	Yes	No	Yes	No	Yes
Demo. Char.	No	Yes	No	Yes	No	Yes
N	1 643	1 640	1 644	1 642	1 644	1 642

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	0.822	0.803	0.718

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 a sample without 25 % of respondents who spent the least time on the corresponding task.

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

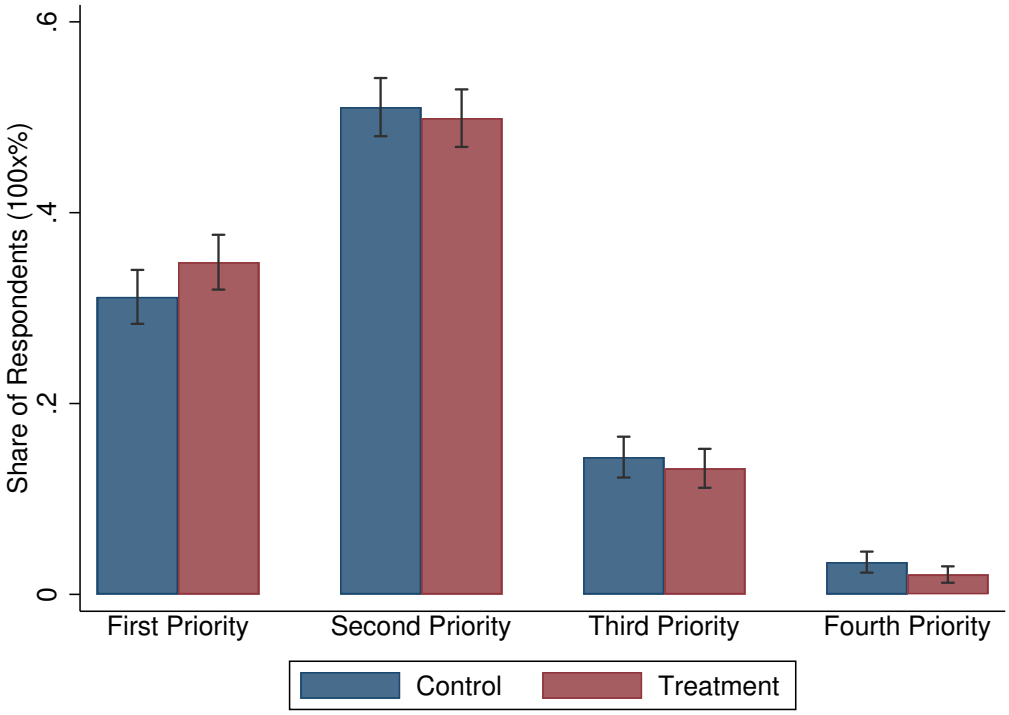
system is likely affected by the environment of the survey experiment itself (through for example the experimenter demand effect) and thus it is barely generalizable, it is a good signal of a relevance of the issue. Figure 4 shows shares of respondents who ranked fairness of the judicial system as the first, the second, the third 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. In the treatment group, 34.8 % of respondents ranked fairness of judicial system as the top priority, while in the control group it was 31.2 %.

Information about sentencing disparity increases willingness to sign a petition. Almost two thirds of respondents declared their interest in reading and signing the petition. After reading the petition, slightly more than 25 % of all respondents, including those who were initially dropped from the *baseline* sample, provided their email addresses to have the signature sheet sent. Figure 5 shows that the share of respondents interested in petition is higher in the treatment group (65.6 %) than among respondents in the control group (61.3 %). In the similar vein, the two bars on the right show that while 23.6 % of respondents randomized into control group requested the signature sheets, in the treated group it was 26.8 %. Since I do not have the individual level data about the latter measure and thus do not observe time spent on the treatment and the control slides, I rely on a comparison at the randomization level. Using different samples arguably contributes to the large gap of almost 40 percentage points between those interested in reading and signing the petition and the numbers of emailed addresses provided.

To test the treatment effect formally, I: (i) run univariate OLS; (ii) extend the OLS regression by controlling for prior attitude and demographic variables; (iii) simulate the exact p-value; and (iv) for the rank of priorities run the Wilcoxon rank-sum test. I use two different samples: (i) the *baseline* sample; and (ii) samples that drop 25 % of respondents with the least time spent on the corresponding task. As a variable of interest, I use three different dummy variables: *Petition*, an indicator of declared interest in reading and signing the petition; (ii) *Top Priority JS* which equals 1 if fairness of the judicial system was the first priority; and (iii) *High Priority JS* which equals 1 if fairness of the judicial system was ranked either as the first or the second priority.

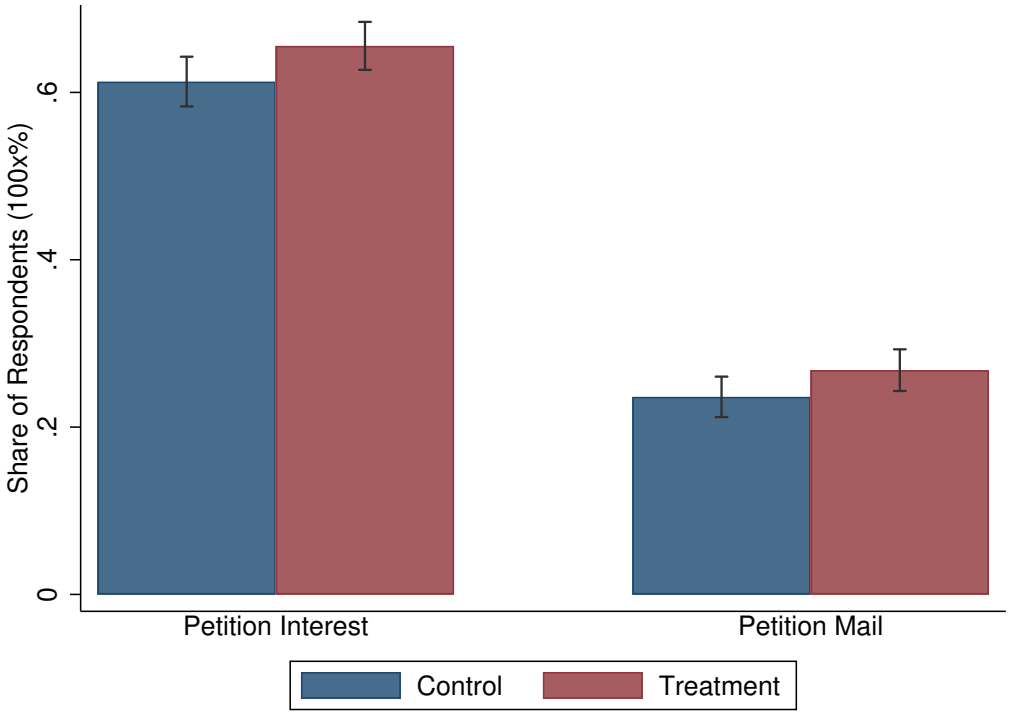
Panel A of table 3 shows results based on the *baseline* sample. Information about sentencing disparity increased willingness to read and sign a petition by 4.3 percentage points (7 %) and the effect is statistically significant at the 5 percent significance level. In a similar vein, the treatment increases *Top Priority JS* by 3.6 percentage points (11.5 %), which is marginally statistically significant; and *High Priority JS* by 2.5 percentage points (TODO in %), which is not statistically significant at the conventional levels. Panel B reports results from the restricted samples. Restricting the dataset leads to

Figure 4: 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 issue that were to ranked to were sufficient highway infrastructure, safety in the Czech republic, high-quality teachers in education system.

Figure 5: Interest in Signing Petition.



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

Table 3: Treatment Effect on Political Preferences

Panel A: Full Sample

	Petition		Top Priority JS		High Priority JS	
Treatment	0.043**	0.043**	0.036*	0.038*	0.025	0.026
	(0.021)	(0.021)	(0.021)	(0.020)	(0.016)	(0.016)
Prior Attitude	No	Yes	No	Yes	No	Yes
Demo. Char.	No	Yes	No	Yes	No	Yes
N	2 096	2 093	2 096	2 093	2 096	2 093

Panel B: Restricted Sample

	Petition		Top Priority JS		High Priority JS	
Treatment	0.057**	0.058**	0.046**	0.048**	0.026	0.028
	(0.023)	(0.023)	(0.023)	(0.023)	(0.018)	(0.018)
Prior Attitude	No	Yes	No	Yes	No	Yes
Demo. Char.	No	Yes	No	Yes	No	Yes
N	1 670	1 667	1 629	1 626	1 629	1 626

Standard errors in parentheses

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

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

	Petition	Top Priority JS	High Priority JS
Exact p-value	0.043**	0.079*	0.127
		Rank Judicial System	
Wilcoxon test		0.036**	

Panel A shows results from univariate and multivariate OLS regressions of *Treatment* on dummy variables that measure whether one (i) is interested in reading a petition; (ii) ranks fairness of judicial system as the top priority; (iii) ranks fairness of judicial system as the top or second priority.

Panel B shows results from univariate and multivariate OLS regressions of *Treatment* on dummy variables that measure whether one (i) is interested in reading a petition; (ii) ranks fairness of judicial system as the top priority; (iii) ranks fairness of judicial system as the top or second priority using a sample without 25 % of respondents who spent the least time on the corresponding task.

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

an increase of the point estimates for all six specifications. In particular, respondents are more likely to be interested in signing petition by 5.7 percentage points (8.8 %) and *Top Priority JS* increased by 4.6 percentage points (TODO in %). The effect on *High Priority JS* increases, but remain statistically insignificant.

Panel C provides exact p-values of the treatment effect derived on the *baseline* sample and confirms changes in policy preferences evoked by the treatment information are statistically significant. Finally, Wilcoxon rank-sum test also suggests that the effect caused by the treatment on the rank of fairness of the judicial system among the priorities is statistically significant. [17]

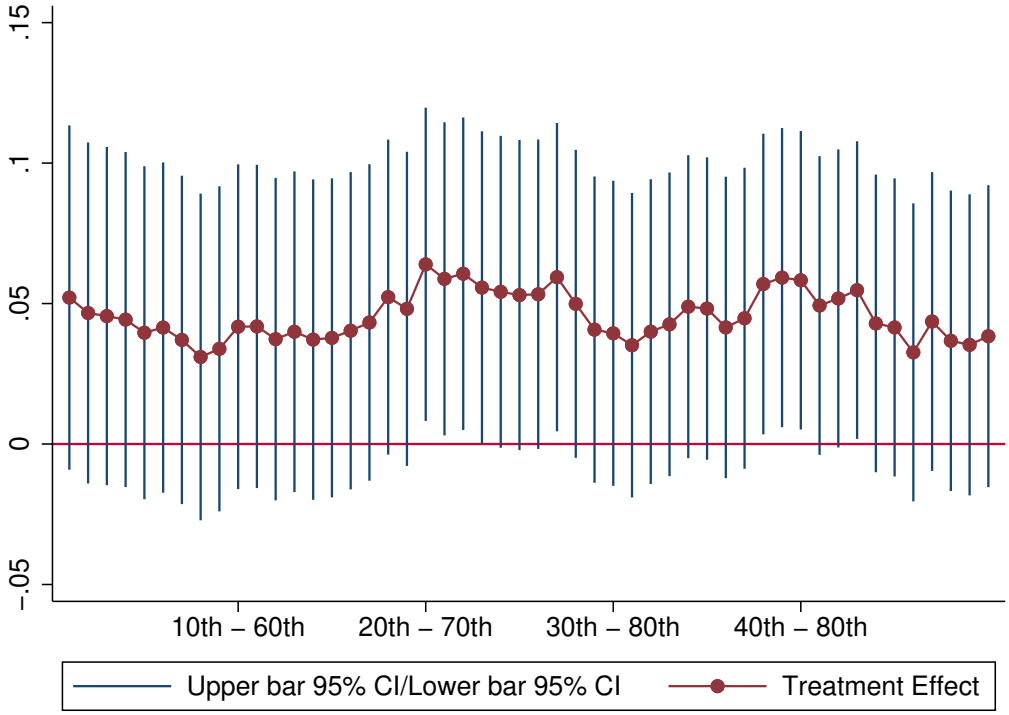
I next provide an additional insight about attention sensitivity of the treatment information on the willingness to read and sign the petition. A comparison of panel A and panel B reveals that as I drop 25 % observations that spent the least time on the petition slide, the point estimates increase. Intuitively, quick responses are likely less considered and thus less informative about the treatment effect. To understand how sensitive the results are, I re-estimate a sequence of treatment effects on a sample with respondents with a different level of attention. Starting with a window of 50 % observations that spent the least time on the petition task, I estimate a univariate regression for *Petition*. Then I move the window by 1 percentage points and re-estimate the same regression on a slightly different sample. I repeat the process 50 times, until the final regression is not estimated on a sample of those who spent the most time on the slide. Figure 6 plots the point estimates with their 95% confidence intervals. Since each of the regression is estimated with half of the observations than the main results presented in panel A of table 3, the confidence intervals are wider. Importantly, the figure shows that the treatment effect does not seem to be systematically affected by the time respondents spent on the task much, the point estimates of the treatment effect fluctuates around the 5 percentage points.

Finally, I present a comparison of shares of respondents who provided their email addresses (*Petition Mail*) between the control and the treatment groups. Disadvantage of this measure is that is based only on aggregate data and it is impossible to focus solely on respondents who paid enough attention to the treatment and the control slides and to control for their observed characteristics. Therefore, I report results only from: (i) a univariate OLS regression; and (ii) simulated exact p-value.

First row of table 4 shows results from the univariate OLS regression and suggests that even on the aggregate level the treatment effect is marginally statistically significant. Information about sentencing disparity increases respondents' willingness to be involved

[17] Simulated exact p-value and Wilcoxon rank-sum test also suggest that the effects estimated on the restricted sample are statistical significant at lower significance levels.

Figure 6: Sensitivity of Treatment Effect with Varying Attention



Notes: The figure shows treatment effects on willingness to read and sign the petition with the 95 % confidence intervals for different groups by time spent on the petition task. The estimate on the left corresponds to the sample of 50% of respondents who spent the least time, while the estimates on the right corresponds to the sample of 50 % of respondents who spent the most time on the task. Compared to the *baseline* sample, the standard errors are by almost 0.5 percentage points larger and confidence intervals by 2 percentage points wider.

Table 4: Treatment Effect on Share of Email Addresses Provided

	Petition Mail
Randomization Status	0.032*
	(0.018)
Exact p-value	0.071*

Standard errors in parentheses

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

in policy debate and to provide me with email address to have the petition sent by 3 percentage points, which represents an increase by 13.5 %. The second row of table 4 presents the exact p-value of 0.07 and confirms that the treatment effect is marginally significant.

3.2.5 Heterogeneous Treatment Effect

I move beyond the average treatment effect and provide evidence that the treatment effect is driven mostly by: (i) female respondents with children; and (ii) respondents who reported low level of approval of the judicial system a priori. Both drivers of the effect are intuitive. The former comes from the particular type of crime used in the treatment slide – *failure to pay alimony* – to which female respondents with children are arguably more sensitive than any other sociodemographic group. Since the differences between coefficients for different subsamples are not statistically significant, I view these results as suggestive, but informative about the underlying narrative.

Table 5: Heterogenous Treatment Effect

Panel A: *Petition* w.r.t Gender and Parenthood

Treatment	0.008	0.073**	0.061**	-0.004	-0.033	0.018	0.106***	0.002
	(0.031)	(0.029)	(0.025)	(0.040)	(0.062)	(0.054)	(0.032)	(0.038)
Female	×	✓	All	All	✓	×	✓	×
# Kids > 0	All	All	✓	×	×	×	✓	✓
N	996	1 100	1 508	588	251	337	849	659

Panel B: *High Priority JS* w.r.t Gender and Parenthood

Treatment	-0.001	0.047**	0.021	0.035	0.050	0.018	0.046*	-0.012
	(0.024)	(0.022)	(0.019)	(0.031)	(0.043)	(0.044)	(0.025)	(0.029)
Female	×	✓	All	All	✓	×	✓	×
# Kids > 0	All	All	✓	×	×	×	✓	✓
N	996	1 100	1 508	588	251	337	849	659

Panel C: *Petition* w.r.t Prior Attitude

Treatment	-0.002	0.091***	0.038*	0.102	-0.006	0.043	0.082***	0.239*
	(0.029)	(0.030)	(0.022)	(0.075)	(0.031)	(0.091)	(0.031)	(0.131)
High Approval	✓	×	All	All	✓	✓	×	×
High Perception SD	All	All	✓	×	✓	×	✓	×
N	1 085	1 011	1 919	177	963	122	956	55

Standard errors in parentheses

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

High Approval equals 1 if a respondent reports JD works well (definitely yes or rather yes); 0 otherwise

High Perception SD equals 1 if a respondent perceives sentencing hetorogenienty (definitely and rather yes); 0 otherwise

To demonstrate the heterogenity of the effects with respect to gender and parenthood, I split the *baseline* sample along these two characteristics: respondents' gender and whether they have at least one child. I report results on willingness to read and sign the petition (*Petition*), as presented in panel A and results on dummy indicating that the fairness of judicial system is among the top two priorities ($High\,Priority\,JS$) in panel B of table 5. In particular, I report results form a simple univariate OLS regressions on each of the corresponding subsample. (TODO - TopPriorityJS)

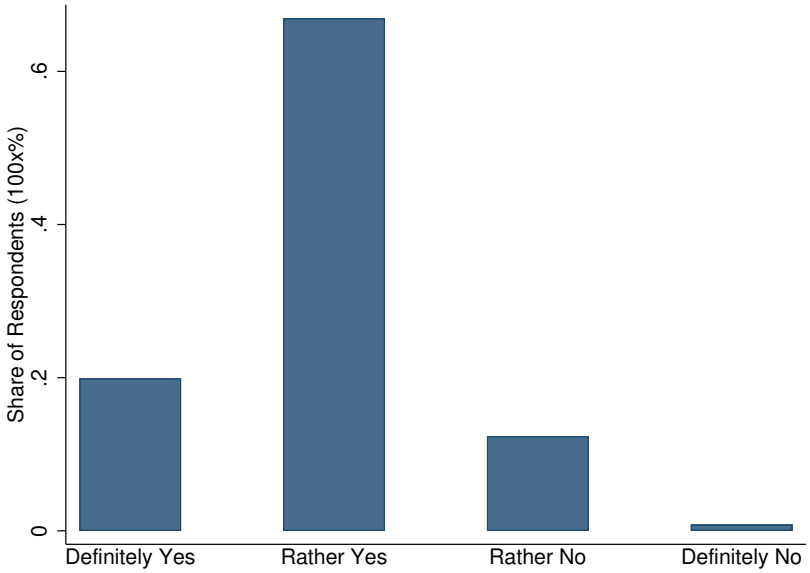
Consider first *Petition* as the outcome of interest. Among female respondents, the treatment effect is of 7.3 percentage points, which corresponds to an increase by 11.2 %; effect estimated on a sample of respondents with at least one child, the effect is 6.1 percentage points (9.3 %); and finally, the effect estimated on among female respondents with at least one child is 10.6 percentage points, which represents an increase by 15.8 %. Consistently with the effects *Petition*, also the effect on $High\ Priority\ JS$ suggests the same demographic groups are sensitive to the treatment information. Note, however, that the effect on $High\,Priority\,JS$ is less prominent than the one on *Petition*. Overall, female respondents with parenthood experience tend to react to the information about sentencing disparity in the crime of *failure to pay alimony* more than other groups. The effect is economically sizeable and statistically significant.

To study heterogeneity with respect to the prior level of approval of the judicial system and the prior perception of sentencing disparity, I divide both measures into two groups; high and low level of reported approval and high and low level of perceived sentencing disparity and report results from a univariate OLS regression with *Petition* as the main outcome of interest. Panel C of table 5 shows the results. The second column shows that among respondents who reported a low level of approval of the judicial system, the treatment increases their willingness to read and sign the petition by 9.1 percentage points, which represents an increase of 14 %. Additionally, the last two columns provide additional evidence supporting that the treatment effect is particular strong among those respondents who reported a low approval rate.

4 Discussion

I next discuss the interpretation of the treatment and add thoughts on external validity and generalizability of the results for cases in which the information about sentencing disparity would become widespread news.

Figure 7: Perception of Credibility of Information Provided



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

4.1 Interpretation of the treatment effect

An important concern in studying trust and trustworthiness based on information treatment is whether respondents trust the information provided. At the beginning of the experiment, all respondents were notified that the information provided to them are truthful and based on data from the Ministry of Justice. This may evoke tension between implicitly asking respondents to trust the information based on data provided by the Ministry of Justice and, at the same time, ask them whether they trust the judicial system. To understand the degree of potential risk of mistrusting the information provided, I elicited directly the perceived credibility of the information. To limit effects of asking the question regarding the credibility of the information provided on respondents' actions and decisions in the experiment, I asked at the very end of the experiment. 87 % of respondents declared that they view the data credible and only less than 1 % of them selected an extreme choice of *definitely not credible*. This suggests that the results are not systematically affected by respondents' mistrust in the information treatment. See figure 7.

For the interpretation of the results, it is important to understand how respondents understand and interpret the information provided and what difference in the perception is caused by the treatment. The core of the treatment is to provide information about varying levels of sentencing disparity, i.e. the second moment of the distribution

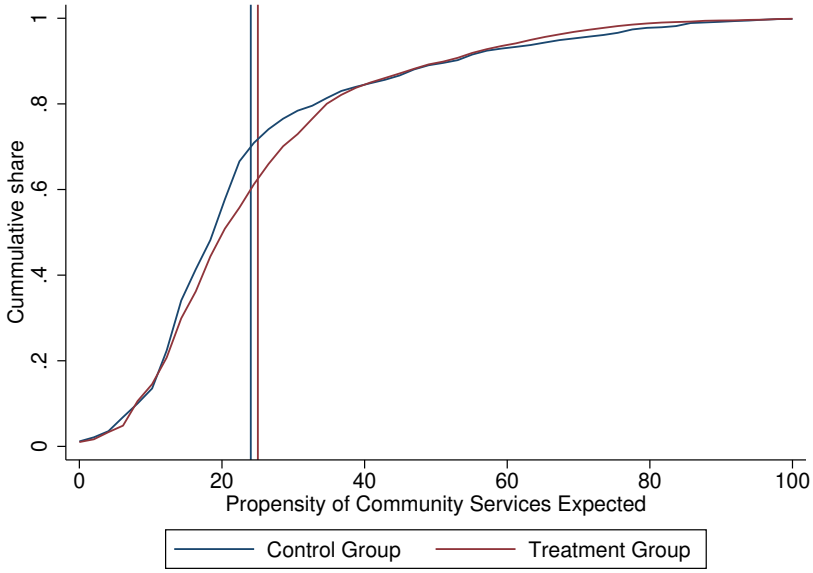
of sentencing decisions. This is a novel feature, as most of the information and survey experiments existing in the literature exogenously vary beliefs about the first moment of a relevant distribution, e.g. a probability of audit, a share of high-skilled immigrants. Varying signals about sentencing disparity is challenging, as it may be more complicated 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 services for *failure to pay alimony*, i.e. the first moment of the distribution, were numerically identical (16 %). Nevertheless, it is still plausible that the information provided affects respondents' perception of propensity to sentence to community services systematically differently in the control and the treatment group.

To understand whether the threat, I elicited the respondents' expectation regarding the average propensity to sentenced to community services, 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 the community service. On average, respondents in both groups overestimate the true shares. While the national average corresponds to the one of the presented cases, i.e. 16 %, the respondents in the control group expected 24.04 % and in the treatment groups is 25.03 %. The difference between the groups is not statistically significant (p-value = 0.204). Figure 8 shows empirical cumulative distribution functions of expected share of cases sentenced to the 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 suggest that the information provided does not affect the perception of the propensity of sentences to community services systematically differently in the control and the treatment groups.

4.2 Generalizability and Implications

The core of the treatment is a piece of information about sentencing disparity presented in a way in which the news would likely provide such information. Information displayed on the treatment and control slides consist of an informative headline that summarizes the main message, a self-explaining figure that represents the core of the message and a brief interpretation of what it represents. However, in other important dimensions the treatment is far from what one would expect to observe if the information about sentencing disparity attract attention of politicians, media, and society and become a mainstream topic. In particular, the information provided in the experiment is one-time, brief, and isolated. On the contrary, if the same information makes it in the

Figure 8: Expected Share of Cases Sentenced to Community Service



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

headline news, the information would be widespread, discussed on spotlight repeatedly, and likely becomes a topic among peers and colleagues. All three aspects are likely to magnify the awareness of citizens and potentially increases the treatment effect. It would be, for example, more complicated not to pay attention to that information. Therefore, compared to a hypothetical situation in which the information about sentencing disparity would make headlines, then the treatment effects estimated in this project are barely easily generalizable.

Evidence about heterogeneity of the treatment effect and specifically the fact that the treatment effect seems to be driven by two groups: female respondents with children and respondents with a priori low level of approval of the judicial system are informative for a discussion about external validity. These two groups are arguably triggered by different mechanisms. While the former group – *topic-sensitive* respondents – is sensitive to the specific information provided and may not be that sensitive to sentencing disparity in different types of crime (or information about performance of different public institutions), the latter group – *generally concerned* respondents – is arguably someone who does not approve the judicial system and information about sentencing disparity, regardless of type of crimes, leads to an increased will to sign the petition. Therefore, information about sentencing disparities in different types of crimes has the potential to affect different *topic-sensitive* respondents, while the *generally concerned* respondents are potentially sensitise

to the information about sentencing disparity in general. This suggests that information about sentencing disparity in more types of crime would lead to more sizeable effect through an extensive margin channel, as more sociodemographic groups would react. I view this as an additional channel through which information about sentencing disparities in all (other) types of crimes may increase the estimated effects.

If both mechanisms were to boost the treatment effect and causes distrust and/or restrict reliance on the judicial system, my results are still informative about the relative magnitude of the effect and where it hits the most. Taking the effects at their face values suggests that the general public would first and most strongly react in terms of policy preferences, including a (active) participation in the policy discussion, whereas distrust and intention to avoid and substitute the judicial system, if appear, would come afterwards and feebly.

My results have also implications for a discussion about publicly available information about a performance of state institutions. Suppose that a public authority is responsible for a given institution and he decides on making information about the performance publicly available (e.g., public procurement contracts of contacting authorities). If he suspects that the information may hurt him, his reputation, and his future in the office, he may, in order to keep the information confidential, argue that if the information were public, it would cause a distrust with a significant (social) cost. In the perspective of my results, however, the general public is likely to demand policy changes which may jeopardize the positions of incumbent public authority first. The distrust, if were to appear, would follow afterwards.

Concluding Remark

Many decisions of public authorities around the world feature some level of discretion that often leads to disparity in outcomes, including disparities in sentencing. The fact that the sentence of offenders depends on a personality of the judge assigned to the case contradicts rules of modern democracies. Importantly, if the sentencing disparities were to undermine the perceived trust and affect citizens' behavior, it may caused an increase of cost of social interactions and thus reduce economic welfare.

To understand consequences of sentencing disparity on trust in the judicial system and relevant behavior, I conduct a survey experiment in which I exogenously vary information about sentencing disparity among judges within a regional court in the Czech republic. In particular, respondents were informed about share of cases *failure to pay alimony* – one of the most frequent type of crimes – in which the convicted offender was sentenced to a punishment of community service. Respondents in the treatment groups were presented

with a court in which judges differ in their propensities to sentence to community service; whereas the active control group was informed about a court in which the differences in shares of cases in which the convicted offender was sentenced to community service are negligible. I study three groups of outcomes: (i) declared institutional trust, (ii) reliance on the judicial system; (iii) policy preferences.

I find that the treatment increases a willingness to read and sign petition that invites politicians to particularize sentencing principles and shifts a view on fairness of judicial system which is considered a more important policy priority. The effect is driven by two types of respondents: (i) female respondents with children; (ii) respondents who ex-ante do not approve how the judicial system works. The former group is arguably more sensitive to the particular type of crime in the information treatment – *failure to pay alimony* than any other demographic groups. The heterogenity of the effect is informative about generalization, as information about sentencing disparity in other types of crime is more likely to evoke different type of respondents to react.

I find no evidence that the information treatment affects declared institutional trust neither in the judicial system, nor in other relevant institutions such as the police. Respondents in both the control and the treatment groups report relatively high level of trust in the judicial system. Probably more importantly, the treatment does not change (intended) willingness to rely on the judicial system. Overwhelming majority of respondents from both groups recommend a single mother to apply to court in a situation she hesitates. Additionally, I find no evidence that the information treatment increases interest in learning about substitute to the formal judicial system.

References

Acemoglu, D., Cheema, A., Khwaja, A. I., and Robinson, J. A. (2020). Trust in state and non-state actors: Evidence from dispute resolution in pakistan.

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.

Alesina, A., Miano, A., and Stantcheva, S. (2018). Immigration and Redistribution.

Athey, S. and Imbens, G. W. (2017). The econometrics of randomized experiments. In *Handbook of economic field experiments*, volume 1, pages 73–140. Elsevier.

Beaudry, P. and Willems, T. (2018). On the Macroeconomic Consequences of Over-Optimism. *IMF Working Papers*, 18(122):1.

Bérgolo, M. L., Ceni, R., Cruces, G., Giaccobasso, M., and Perez-Truglia, R. (2019). Tax Audits as Scarecrows: Evidence from a Large-Scale Field Experiment. *NBER Working Papers*.

Bhuller, M., Dahl, G. B., Løken, K. V., and Mogstad, M. (2020). Incarceration, recidivism, and employment. *Journal of Political Economy*, 128(4):1269–1324.

Bursztyn, L., Egorov, G., and Fiorin, S. (2019). From extreme to mainstream: The erosion of social norms.

Cappelen, A., Haaland, I., and Tungodden, B. (2019). Beliefs about Behavioral Responses to Taxation.

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*.

Chen, D. L. and Philippe, A. (2017). Clash of norms: Judicial leniency on defendant birthdays.". *Institute for Advanced Study in Toulouse/Toulouse School of Economics*.

Cruz, C., Keefer, P., and Labonne, J. (2018a). Buying informed voters: New effects of information on voters and candidates.

Cruz, C., Keefer, P., Labonne, J., and Trebbi, F. (2018b). Making policies matter: Voter responses to campaign promises. Technical report, National Bureau of Economic Research.

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

Di Tella, R. and Schargrodsky, E. (2013). Criminal recidivism after prison and electronic monitoring. *Journal of Political Economy*, 121(1):28–73.

Dobbie, W., Goldin, J., and Yang, C. S. (2018). The Effects of Pretrial Detention on Conviction, Future Crime, and Employment: Evidence from Randomly Assigned Judges. *American Economic Review*, 108(2):201–240.

Eurobarometer, S. (2018). Public opinion in the european union.

Gaulé, P. (2018). Patents and the Success of Venture-Capital Backed Startups: Using Examiner Assignment to Estimate Causal Effects. *Journal of Industrial Economics*, 66(2):350–376.

Goncalves, F. and Mello, S. (2017). Does the Punishment Fit the Crime? Speeding Fines and Recidivism. *Ssrn*.

Grigorieff, A., Roth, C., and Ubfal, D. (2016). Does Information Change Attitudes Towards Immigrants? Representative Evidence from Survey Experiments.

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.

Haaland, I. and Roth, C. (2017). Labor Market Concerns and Support for Immigration.

Haaland, I. and Roth, C. (2018). Beliefs about Racial Discrimination: Representative Evidence.

Haaland, I., Roth, C., Wohlfart, J., et al. (2020). Designing information provision experiments. Technical report, University of Copenhagen. Department of Economics. The Center for Economic ....

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.

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

Kuziemko, I., Norton, M. I., Saez, E., and Stantcheva, S. (2015). How elastic are preferences for redistribution? evidence from randomized survey experiments. *American Economic Review*, 105(4):1478–1508.

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.

Mueller-Smith, M. (2015). The criminal and labor market impacts of incarceration. *Unpublished Working Paper*, 18.

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.

Mustard, D. B. (2001). Racial, ethnic, and gender disparities in sentencing: Evidence from the us federal courts. *The Journal of Law and Economics*, 44(1):285–314.

OECD (2017). Trust and Public Policy: How Better Governance Can Help Rebuild Public Trust.

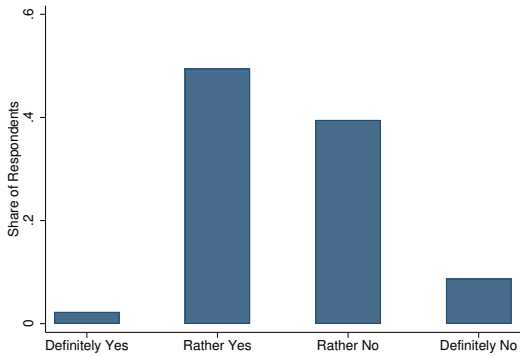
Sampat, B. and Williams, H. L. (2019). How do patents affect follow-on innovation? evidence from the human genome. *American Economic Review*, 109(1):203–36.

Sporer, S. L. and Goodman-Delahunty, J. (2009). Disparities in sentencing decisions. *Social psychology of punishment of crime*, pages 379–401.

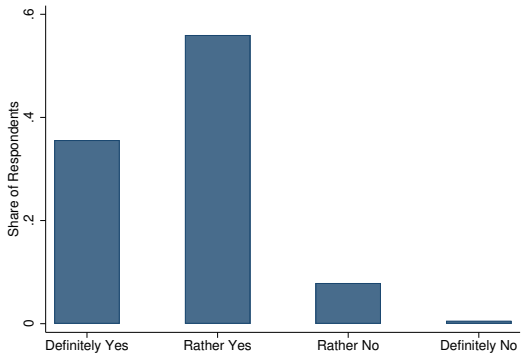
Steffensmeier, D. and Demuth, S. (2006). Does gender modify the effects of race–ethnicity on criminal sanctioning? sentences for male and female white, black, and hispanic defendants. *Journal of Quantitative Criminology*, 22(3):241–261.

Appendix

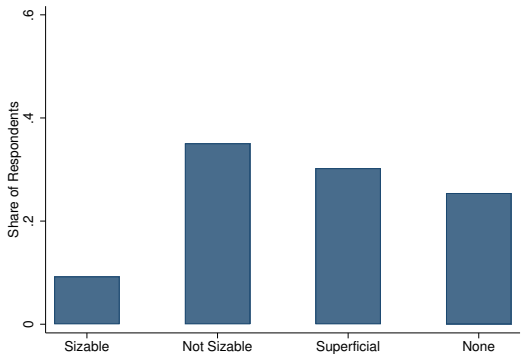
Figure 9: Prior Attitude towards Judicial System



(a) Prior Approval of the JS



(b) Perception of Sentencing Disparity



(c) Prior Experience with the JS

(a) Juridical system works well in the Czech Republic. (b) Depending on judge, similar cases can be sentenced differently. (c) Considering how often you or people you know come into contact with juridical system, what experience you think you have?

Table 6: 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	0.340
Rather yes	0.488	0.501	0.567
Rather no	0.392	0.398	0.771
Definitely no	0.094	0.081	0.311
Perception of Sentencing Disparity			
Definitely yes	0.372	0.341	0.138
Rather yes	0.547	0.572	0.261
Rather no	0.075	0.082	0.564
Definitely no	0.006	0.006	0.968
Experience with the Judicial System			
Sizable	0.087	0.098	0.375
Not sizeable	0.354	0.347	0.734
Superficial	0.304	0.301	0.877
None	0.255	0.254	0.956
Demographic Characteristics			
Education			
University	0.249	0.255	0.765
Highschool	0.697	0.692	0.825
Elementary	0.054	0.053	0.901
Marital Status			
Single	0.210	0.187	0.175
Cohabitation	0.171	0.182	0.466
Married	0.442	0.423	0.369
Divorced	0.150	0.167	0.277
Widowed	0.027	0.041	0.087
Male	0.485	0.466	0.400
Age	44.71	44.90	0.760
At least 1 child	0.717	0.722	0.818
Number of children	1.435	1.490	0.295
N	1,036	1,060	

Online Appendix

Script

Throughout the experiment, the respondents are shown x or y slides, depending on their choices. Furthermore, at the end of the survey experiment they can choose to go to a website with petition to sign. The survey experiment consists of three blocks prior-treatment questions, treatment slides, post-treatment questions.

Prior-treatment Questions

Prior 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. *This survey is about how courts work. All information provided are truthful and based on data from the Ministry of Justice.*

Slide 2. *We are going to show you two statements and ask you how much you agree with that statements.*

Slide 3. *Sentencing decisions depend on the judge assigned. Depending on judge, similar cases can be sentenced differently.*

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

Slide 4. *Judicial system works well in the Czech Republic.*

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

Slide 5. *Considering how often you or people you know well come into contact with judicial system, what experience you think you have?*

- *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 slide consist of a figure of judges and shares of cases sentenced to community services and a brief explanation of what the figure represents. Prior the control and treatment slide, I inform them about what I they are about to see.

Slide 6. *One of the most frequent crimes in the Czech republic is failure to pay alimony. In last 3 years, courts in the Czech Republic sentenced more than 13,000 cases. The punishment can be probation period, incarceration, community service, and 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 the community services as their main punishment.*
Judges sentence very similarly.
Regardless what judge is assigned to the case, the offender has very similar probability of being sentenced to community services.

Slide (Treatment Group)**.** *At this regional court, 16 % of the convicted are sentenced to the community services as their main punishment.*
Judges sentence differently.
Judge C sentence to the community service a third of his/her cases
Other judges sentence to the community service less than 10 % cases, instead they choose different types of punishments
Cases are assigned to judges at random: an offender assigned to the judge C has three time higher probability of being sentenced to community services compared to being 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 is offender sentenced to the community services as the main punishment for failure to pay alimony?*

Slide 8. *How much do you trust in the following institutions ...* {*judicial system; police; 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 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 (the arbitration) that has several advantages compared to the juridical system.*

- *Want to know more*
- *Not interested*

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

Slide 11. *We would like to ask your advice. Your response can be used by NGO (vasevyzivne.cz), as information for their clients. Please read the following story: Jane has two kids and their father does not pay (alimony/child support). She wonder whether she should 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 sing a petition that invites politicians to suggest particularizing sentencing principles? Such principles would assist judges in 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 before the petition promised/and very last thing), 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 representation to particulates 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. Should you be interested in signing this petition, leave us with your email address and we will send you 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 support xx of sentencing principles

Sentencing decisions often crucially (TODO)

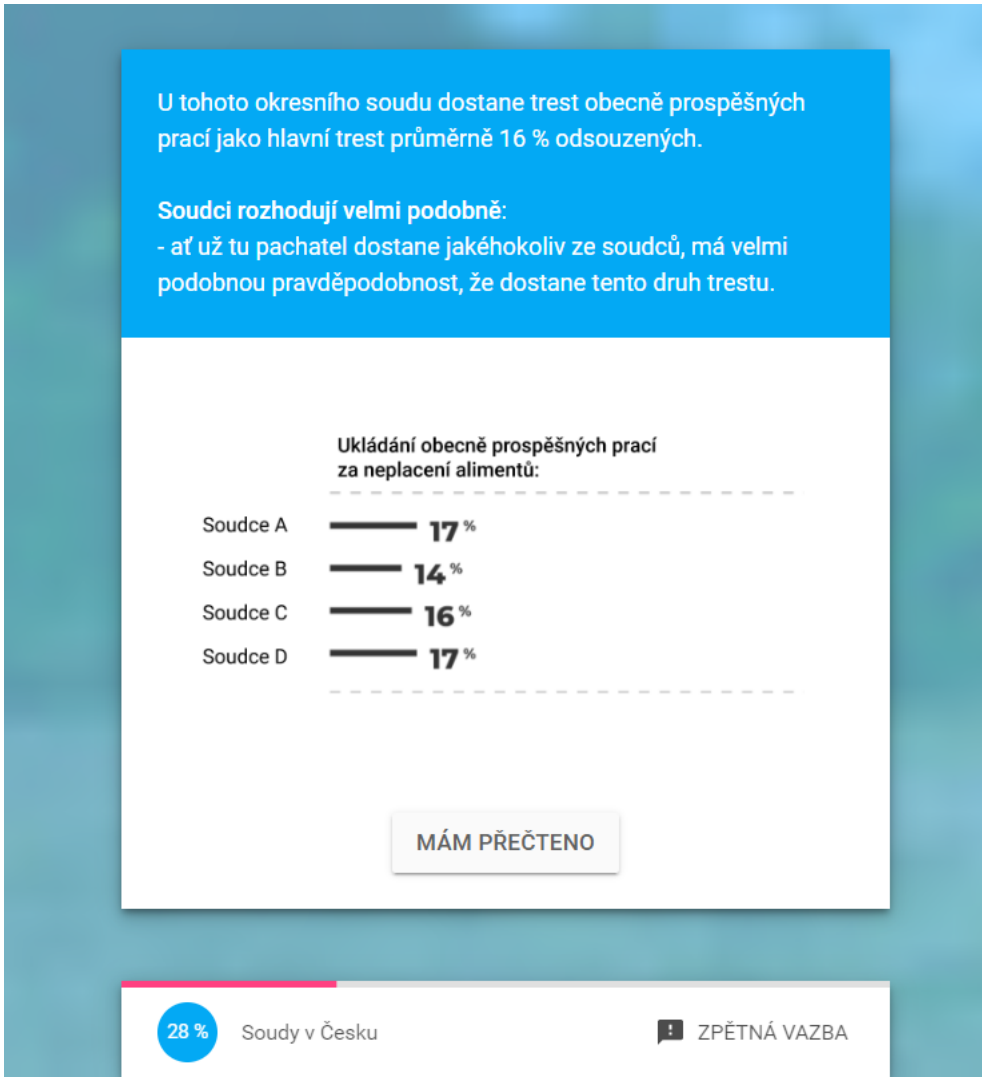
Debriefing Letter

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

The information that was showed to you is real and describe a situation at one of the regional courts. However, the situation may not correspond to the situation at other regional courts. In the Czech republic, there are several dozens of regional courts that may differ in propensity to sentenced 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.

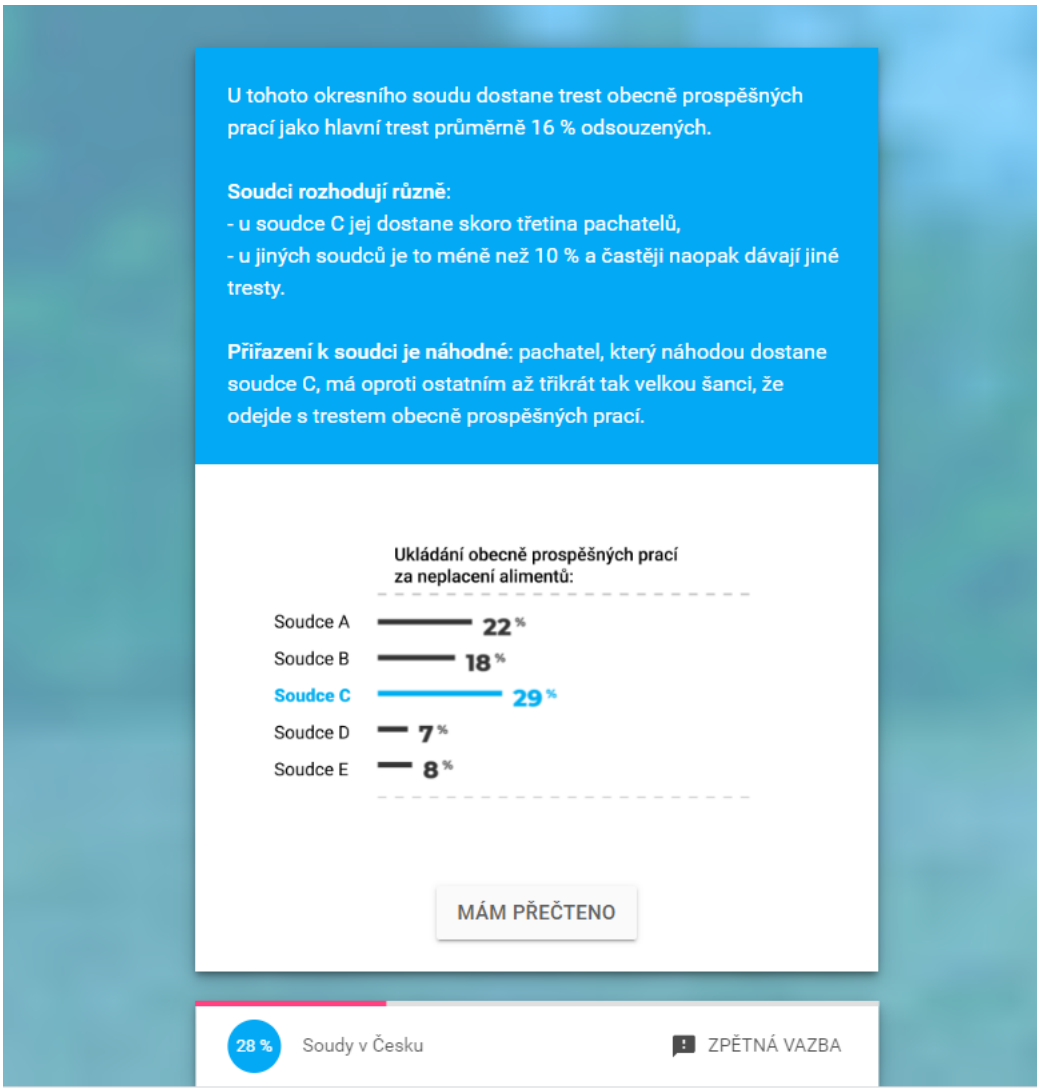
Figure 10: Control Slide



Notes:

Treatment and control slides

Figure 11: Treatment Slide



Notes: