Sentencing Decisions Around Thresholds Dividing Subsections of Offenses: Theory and Experiment*

Jakub Drápal[a] and Michal Šoltés[†a,b]

[a]Department of Economics, Faculty of Law Charles University[‡]
[b]CERGE-EI[§]

Preliminary and incomplete draft, please do not circulate.
The most recent version

February, 2021

Abstract

In order to minimize sentencing inconsistency, offenses are typically divided into subsections, resulting in similarly serious cases being sentenced according to the same sentencing ranges within the same subsection. In many cases, the classification of a case into corresponding subsection depends on exceeding a given threshold. We study to what extent the division of offenses can introduce a new source of sentencing inconsistency. We argue that the effect of thresholds on sentencing decisions can be decomposed into two opposing mechanisms: the severity and the reference effect. We conducted an experiment with Czech prosecutors that provides evidence that thresholds can cause an enormous increase in harshness in sentencing. In the most striking example, theft with damage around 50 000 CZK (approximately $ 2 300), the threshold increases an average length of incarceration by more than half. The experiment further provides suggestive evidence of the proposed mechanisms.

*We thank to Libor Dušek, Ondřej Kubů, Filip Matějka, Josef Montag, Vladimír Novák, Jiří Pavlík, Ondřej Šťastný, Jan Zápal, and the Prosecutor General's Office and to prosecutors who participated in our survey. This study was supported by Charles University, GAUK project No. 218218 and also received funding from the European Research Council under the European Union's Horizon 2020 research and innovation programme (grant agreement No 678081). Preregistration: AEARCTR-0006023

[†]Email address: msoltes@cerge-ei.cz

[‡]Charles University, Faculty of Law, nám. Curieových 901/7, 116 40 Prague, Czech Republic.

[§]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.

1 Introduction

In most legal systems, sentencing policy is governed by two rules. Sentencing decisions are required to be individualized and at the same time the judicial system strive to be consistent. That leads to an extremely complex decision, in which the sentencing decision at the same time respects detailed information of a particular case and takes into account sentences imposed in similar cases by other judges. To enable the fulfillment of the former requirement, wide discretion is offered to judges, who are arguably in the best position to learn and evaluate the detailed information related to the particular case. To limit potential sentencing inconsistency that may arise from different approaches of judges, the space for such discretional decision is narrowed down by either criminal code or sentence orientations.

The most common measure structuring sentencing discretion are sentencing ranges which specify what sentences can be imposed for a given offense. Additionally, since the same offense can vary significantly in its circumstances and seriousness, it seems reasonable to limit the discretional space of judges not only to a particular offense but also according to characteristics of the offense such as seriousness, creating subsections with specific sentencing ranges. To distinguish among differently serious cases of the same offense, criminal codes (and sentencing guidelines) often rely on measurable and quantifiable variables such as caused damage and amount of drug possessed. Generally, the idea of dividing offenses into subsections with different sentencing ranges and thus treat a theft with a damage of, for example, 50 USD differently than a theft of 500,000 USD is likely an useful measure to limit concerns of sentencing inconsistency. However, a finer division of offenses creates a practical issue of how to structure the criminal code and in particular how to determine the thresholds and corresponding sentencing ranges.

Thresholds can potentially lead to an increase in sentencing inconsistency. Consider two cases of theft that marginally differ in the amount of damage. If the damages appear on the opposite sides of a threshold dividing the offense into two subsections with different sentencing ranges, almost identical cases face different sentencing ranges. Furthermore, as the composition cases of that subsections differ, the relative seriousness changes as well. Both of the mechanisms can instigate sentencing inconsistency. Therefore, understanding the consequences of the structure of criminal code and in particular how the thresholds dividing subsections of offenses shape sentencing decisions is of a great importance.

In this paper, we introduce a simple theory of how judges impose sentences and apply the theory to study consequences of thresholds and sentencing ranges on sentencing decisions. In particular, we assume that in order to impose a sentence between the low and the upper bounds of the sentencing range, judges relate the exact position to a

relative seriousness of the particular case within the corresponding subsection of the offense. Applying this theory, we show that thresholds can cause a sizeable difference in sentences of two almost identical cases.

The overall effect that a division of offenses has on the sentencing decision can be further decomposed into two opposing effects: severity effect and reference effect. The proposed theory allows us to describe these effects formally and discuss their properties. Moving a case just above the threshold changes two conditions in sentencing decisions. First, the case is now sentenced according to a *higher* sentencing range. We call it the severity effect and show that it leads to more severe sentences. Second, the case that falls just above the threshold is compared to arguably more serious crimes. As a result, the relative position within the sentencing range is likely to be lower. We call it the reference effect and argue that it tends to decrease the sentence.

To provide empirical evidence, we conduct an online experiment with roughly 200 Czech prosecutors. Each participant of the experiment was presented with two vignettes, each describing a different crime case: drug possession and theft. In both cases, the vignettes were prepared in more specifications varying amount of classifying variables (amount of methamphetamine possessed and caused damage) around thresholds. We use the variation in the classifying variable to estimate causal effects of offense divisions. Additionally, to test for existence of the severity and the reference effects, we use two scenarios that rely on alternative legislation introducing an isolated variation in composition of cases in a offense subsections, or sentencing ranges.

The causal effect of threshold on sentencing decisions are enormous leading to a sizeable sentencing inconsistency. Consider first the case of theft in which we vary the amount of damage in four scenarios around two thresholds of 50,000 CZK and 500,000 CZK. To estimate the effect of 50k thresholds, we compare identical cases with marginally different damages of 48,283 CZK and of 51,283 CZK. The results show that the 50k threshold increases average sentence by more than 10 months which represents 50 % increase. Looking at the effect of the 500k threshold, we compare identical cases with damage of 487,092 CZK and 508,213 CZK and estimate the causal effect to be around 4 months that corresponds to 10 %. Interestingly, an increase in average sentence caused by 50k threshold is not statistically different from the increase caused by increasing damage from 51,283 CZK to 487,092 CZK, i.e. approximately by 440 000 CZK. That puts in a context how large the effect of the 50k threshold is.

Looking at the drug possession case, we document that the effect of the threshold is also of a great magnitude and increases the average sentence by more than 6 months (25 %). The effect is in a stark contrast to a 3 % increase of the amount of drug possessed. We next focus on the severity and the reference effect. The design does not allow us to

estimate the full magnitude of these effects. Nevertheless, we can make two observations. First, the fact that the overall effect is positive suggests that the severity effect dominates the reference effect. Second, the test for the reference effect provides suggestive evidence that the effect is negative as expected.

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

2 Theory

2.1 Absolute and relative nature of defining seriousness

Upon sentencing, specific case with multitude of factors is presented to a judge, who is provided with more or less specific guidance that allows him/her to weight all the factors and determine the final sentence. Sentence selection depends primarily upon the seriousness of the committed offense. The type-specific seriousness is determined by the offense classification that carries sentencing range (instituted in penal code or via sentencing guidelines). Aggravating and mitigating factors then lead judges to arrive at a specific sentence within this sentencing range (sometimes being aided by sentence orientations), reflecting the seriousness of the individual offense. Such process – either formalized or implicit – takes place both in legal systems without any specific guidelines (such as Germany or the Czech Republic) and in those operating guidelines of various sorts (such as England and Wales, federal U.S., or the Netherlands).

Determining offense-specific seriousness via legal classification of a committed offense which is connected with a specific sentencing range indicates the general level of severity of the ensuing sanction. Sentencing range serves as the rudimentary signpost of how severe sanction is expected from sentencers: It considers seriousness in absolute term. Offense-specific seriousness also provides limits to sentencers' discretion: Discretion discourse would describe sentencing ranges as the dough of Dworkin's doughnut.

Aggravating and mitigating factors serve to position the offense within the sentencing range. In doing so, cases within the same sentencing range are usually compared among themselves and ranked according to their seriousness. Seriousness is here considered in relative terms as aggravating or mitigating factors indicate to what extent is the present case more or less serious than other cases of the same offense. The relative nature is best manifested when sentencing orientations (e.g. starting points or normal punishments) are officially set, as in these contexts sentencers are required to position the current case with the reference one. Yet consideration of aggravating and mitigating factors is inherently relative also in systems that do not enact sentence orientations: The terms aggravation and mitigation cannot function without being anchored.

Absolute and relative considerations are likely inherent to human deliberations on achieving justice. Aristotle's proportional equality requires that decisions are made both on what is alike (absolute consideration) and similar treatment of those alike (relative consideration). [citace? Nicomachean Ethics, 1130b-1132b]. In sentencing scholarship this dichotomy is best captured by von Hirsch's conception of cardinal and ordinal proportionality von Hirsch (2017): In setting up sentencing ranges, the legislator must not only order the types of offenses on a scale (ordinal proportionality), but the scale must

be anchored (cardinal proportionality). Our discussion of absolute and relative nature of sentence imposition seems to be thus in line with general line of human reasoning.

2.2 Offenses and thresholds dividing them

Criminal justice system divides offenses and attach sentencing ranges to them in various ways. Offenses are usually divided into subsections according to certain factors and specific sentencing ranges are set for every subsection. These subsections are formulated to provide guidance to sentencers by narrowing down the discretion provided to them. The characteristics distinguishing subsections from each other are of different nature, with some of them being quantifiable e.g., the amount of damages [1], amount of narcotics [2] or a number of days when the victim of an assault was not able to work[3]. Quantified thresholds has existed for centuries: For example, distinction in sentencing ranges by amount stolen are around at least since the 1803 when the Criminal Code of the Austrian Empire set thresholds for thefts at 25 and 300 ducats.[4]

Offense subsections with specific sentencing ranges serve to make decision-making more consistent, even though criminal justice systems differ in the way they limit judicial discretion. Some set only upper boundaries (e.g. France), while others fix both lower and upper boundaries with differently stringent possibilities to impose sentences below the lower boundaries Kert et al. (2015). Yet while providing guidance, they might also introduce new kinds of disparities: Cases that are virtually identical – such as theft of 299 or 300 ducats – are sentenced according to a different sentencing range, resulting in possible highly different sanctions. The relative and absolute nature of sentencing discussed earlier enables us to conceptualize the consequences of sentencing two virtually identical cases according to different subsections and sentencing ranges.

Two things change at such threshold. The sentencing range and the composition of comparable cases within the subsection. Upon stealing one ducat more, the offense is being sentenced according to a more severe sentencing range (the type-specific seriousness is different). We call such a change in sentencing caused by this shift *severity* effect. Furthermore, the case does not any more represent the most serious offense within the

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

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

[3] E.g. in Slovakia (Slovak criminal code s. 123/4) or the Czech Republic (Decision of the Czechoslovak Supreme Code n. Tpjf 24/85 [R 16/1986 tr.])

[4] Section 153 and 159.

subsection, but on contrary it is the most lenient variant according to the quantifiable measure within the subsection. We call it a *reference* effect. Severity and reference mechanisms generally influence the consideration of seriousness and ensuing sanction choice in opposite ways: If the case just crosses the thresholds, more severe sentencing range is applied, yet the case is the least serious (regarding the quantifiable measure).

The effects thresholds have on sentence imposition can be analyzed examining current practice. However, separate analysis of the individual mechanisms requires legislative changes of sentencing ranges and of the composition of subsections. Yet we can formulate hypotheses regarding how the mechanisms are likely to influence sentence choice given a change of either sentencing range or the composition of the subsections of an offense. When sentencing ranges change, but the composition of the offenses remains the same, the severity mechanism occurs, while the reference mechanism does not. Reference effect can be imagined as a judge fitting a distribution of cases arranged according to their seriousness within the provided sentencing range. If a new sentencing range is provided, sentences need to be re-adapted to conform the distribution to the new sentencing range. Thus, if a sentencing range for a specific offense subsection is increased (either the lower or the upper boundary is increased), sentences imposed for such offense are likely to increase on average as well.[5]

If only the composition of offense subsections changes – more or less severe cases are added or removed – positions of specific cases are transposed within the offense subsection. If the change in the composition points in one direction, the effect of reference mechanism should be straightforward. To illustrate it, if new cases, which are less serious than the current ones, are added to an offense subsection, sentences for the current ones should increase since all current cases suddenly rank higher within the offense subsection than they did prior to the inclusion of new cases.

Three scenarios are imaginable: Upon crossing the threshold, judges either impose very similar sentences (i) or they impose more severe (ii) or more lenient (iii) sanctions. If the first is true, the requirement of equality to consider alike cases alike is upheld and the thresholds do not introduce any unwarranted disparities. If the second or third is true, the thresholds lead to unjust sentencing: Either a very small increase in seriousness leads to a large increase in severity or perverse sentencing takes place, meaning that more serious offenses are punished leniently.

While we expect the mechanisms to influence all cases within a subsection in the same direction, we believe the size of their effect will not be homogeneous. First, we focus on severity mechanism. Judges' beliefs might strongly counter its effect when they are not

[5]These increases do not have to be the same for all offenses, given that judges are highly likely considering ordinal proportionality across all offenses; the reference mechanism operates likely at inter-offense level as well.

in line with sentencing ranges' changes. For example Italian judges counter the punitive tendencies of the legislator by imposing sentences close to the lower end of the sentencing range Corda (2016). We hypothesize that the more the judges' cardinal perceptions of appropriate sentences will be out of touch with the sentencing ranges, the weaker the severity mechanism will be. Increasing sentencing range for simple driving under influence from 0-0.5 to 0-1 years will thus likely lead to higher sentences, while increasing already high sentencing ranges, such as 0-3 years for simple driving under influence to 0-5 will probably have very small or none effect. Anecdotal evidence is provided by Slovenia adopting a life sentence in 2008, which was never passed since judges thought it to be too severe a sanction Filipčič (2019). We would further expect an increase of lower boundaries to have stronger effect than the increase of higher boundaries: This is also likely since sentences cluster close to the lower boundaries evidenced e.g. in Finland or in the Czech Republic Drápal (2020); Sutela (2020), due to the construction of sentencing ranges; while they have to account for the most serious varieties, the less serious ones are more common: "The statistical distribution of crime severity does not follow the bell curve" (Lappi-Seppälä, 2016, p. 51). We would thus expect an increase from 0-3 to 0.5-3 years to have a larger effect than increasing the sentencing range from 0-3 to 0-4.

Secondly, we expect heterogeneous effects of reference mechanism, since it is by definition about proximity and comparison. Change in offense composition should thus influence the most the cases closest (in the sense of the quantifiable measure) to the added or removed ones, as they are the ones being compared the most often with those added or removed. In other words, the change in the offense composition will influence the most those offenses whose relative position within the subsection changes the most. Thus addition of some cases of bodily harm to the offense of grievous bodily harm will likely influence the "old" least serious cases of GBH, but will have probably none effect on the most serious forms of GBH. Similarly adding the least serious cases of GBH to BH will likely influence the "old" most serious cases of BH, but will likely have none effect on the least serious cases of BH.

2.3 Stylized Framework

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

A sentencing process $\mathcal{S}$ is a two-stages rule that for any offense (x,t) assigns a sentence s. The rule is characterized by a set of thresholds $\tau \in \mathcal{T} = \{\tau^{[0]}, \tau^{[1]}, \tau^{[2]}, \ldots\}$ and

corresponding sentencing ranges $\rho(\tau)$, i.e. intervals $(\rho^-(\tau), \rho^+(\tau))$ that restricts the space for possible sentence.[6] In the first stage, the rule classifies offense (x, t) into an appropriate subsection $\tilde{\tau}$ by comparing the realization of the classifying variable t and the set of thresholds $\mathcal{T}$. In the second stage, the rule suggests a sentence s based on: (i) a sentencing range of the corresponding offense subsection $\rho(\tilde{\tau})$; (ii) a relative seriousness of the case within the corresponding subsection measured by circumstances X and the classifying variable T.

A key object of our framework is a *seriousness function* that assigns a value of its seriousness q to any offense (x, t). A necessary assumption for existence of the *seriousness function* is that sentencers (judges and prosecutors) can compare two different cases (of the same offense) and can decide which one is more serious or that both cases are of the same level of seriousness. In other words, the existence of such a function assumes sentencers' ability to compare and rank cases by its level of seriousness.

Definition 1 (Seriousness Function)**.** A function $q : (X, T) \mapsto \mathbf{R}$ that assigns a seriousness value q to any offense (x, t) is called *seriousness function*.

For further use, we assume that the *seriousness function* satisfies monotonicity condition in both variables. Keeping everything else equal, cases with more severe circumstances and cases with higher classifying variable are assigned a strictly higher value of seriousness. Note, that if the latter were not the case, then it would seem rather odd to divide offenses into subsections using the exact classifying variable.

Assumption 1 (Monotonicity of Seriousness Function)
For any x, t, and any $\varepsilon > 0$ the following holds

$$q(x + \varepsilon, t) - q(x, t) > 0 \tag{A 1.1}$$

$$q(x, t + \varepsilon) - q(x, t) > 0 \tag{A 1.2}$$

Definition 2 (Sentencing Rule)**.** Suppose an offense (x, t) of a given seriousness level $\bar{q}$. Then the sentence assigned to this case is determined by the following two-steps *sentencing rule* $\mathcal{S}$:

$$\tilde{\tau} = min(\tau \in \mathcal{T} | \tau \geq t) \tag{D 2.1}$$

$$s = \rho^-(\tilde{\tau}) + G(x, t; \tilde{\tau}, \bar{q})(\rho^+(\tilde{\tau}) - \rho^-(\tilde{\tau})), \tag{D 2.2}$$

where $G(x, t; \tilde{\tau}, \bar{q}) = \frac{\int_{\tau_{[-1]}}^{\tilde{\tau}} \int_{x_{min}}^{q^{-1}(\bar{q}, l)} f_{X,T}(s,l) \mathrm{d}s \mathrm{d}l}{\int_{\tau_{[-1]}}^{\tilde{\tau}} \int_{x_{min}}^{x_{max}} f_{X,T}(s,l) \mathrm{d}s \mathrm{d}l}$ and $q^{-1}(\bar{q}, t)$ is an inverse function of the

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

seriousness function.

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

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

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

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

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

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

Theorem 1 (Difference in Sentencing)**.** Suppose a set of thresholds $\mathcal{T}$ and two cases of the same offense (x, t) and $(x, t + \epsilon)$, where $\epsilon > 0$, but $\epsilon \to 0$, i.e. two cases that marginally differ in the value of the classifying variable t. Suppose further that $\exists \tau \in \mathcal{T}$ such that $t < \tau < t + \epsilon$.[7] Then,

$$\begin{aligned} &\Delta s = (\mathcal{S}(x, t + \epsilon; \mathcal{T}, \rho(\mathcal{T})) - \mathcal{S}(x, t; \mathcal{T}, \rho(\mathcal{T}))) = \\ &= \underbrace{\Delta\rho^- \left(1 - G(x, t; \tilde{\tau}_1, \bar{q}_1)\right) + \Delta\rho^+ \left(G(x, t; \tilde{\tau}_1, \bar{q}_1)\right)}_{\text{severity effect}} + \underbrace{\Delta G \left(\rho^+(\tilde{\tau}_2) - \rho^-(\tilde{\tau}_2)\right)}_{\text{reference effect}}, \end{aligned} \tag{T 1.1}$$

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

To sketch the proof, let us first define $\tilde{\tau}_1 = min\{\tau | \tau > t\}$ and $\tilde{\tau}_2 = min\{\tau | \tau > t + \epsilon\}$. It then follows that $\tilde{\tau}_2 > \tilde{\tau}_1$. Using (D2.2), the problem simplifies as follows[8]

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

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

[8] See Appendix for more details.

Theorem 1 provides us with a clear prediction of the difference between two cases that marginally differ in the classifying variable around a threshold. The sign of the difference is, however, ambiguous, as there are two likely antagonistic effects: the *severity effect* which is always non-negative[9] and the *reference effect* which can be (and under a likely occurring circumstances will be) negative. If the effects work in opposite directions, then the sign of the overall effect depends on relative strength of these effects. We next discuss the signs of the two effects.

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

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

Corollary 2 (Sign of Reference Effect)
If $\frac{\int_{\tau_1}^{\tau_2}\int_{x_{min}}^{q^{-1}(\bar{q},l+\epsilon)}f_{X,T}(s,l)dsdl}{\int_{\tau_1}^{\tau_2}\int_{x_{min}}^{x_{max}}f_{X,T}(s,l)dsdl}<\frac{\int_{\tau_0}^{\tau_1}\int_{x_{min}}^{q^{-1}(\bar{q},l)}f_{X,T}(s,l)dsdl}{\int_{\tau_0}^{\tau_1}\int_{x_{min}}^{x_{max}}f_{X,T}(s,l)dsdl}$ *then expression* $\Delta G\left(\rho^{+}(\tilde{\tau}_2)-\rho^{-}(\tilde{\tau}_2)\right)$ *– the reference effect – is negative.*

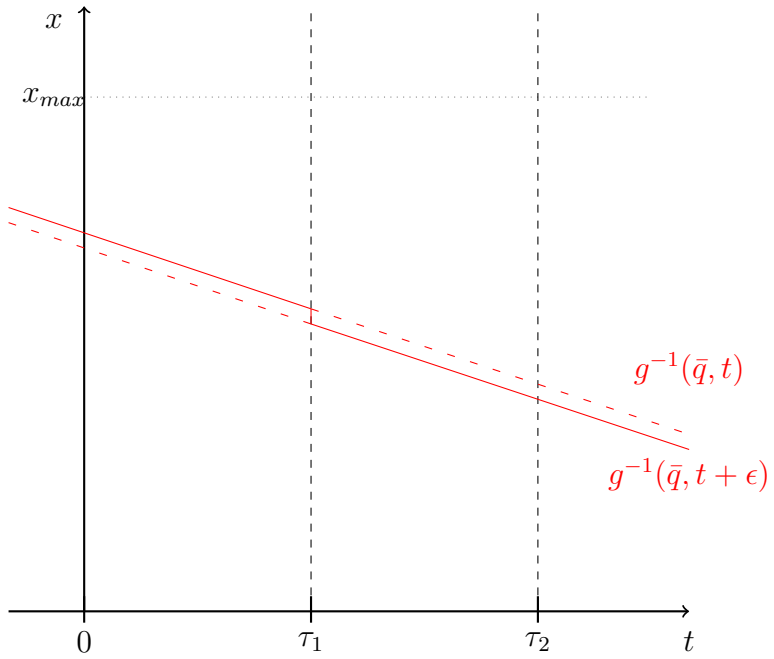
Note that the sign of the *reference effect* corresponds to the sign of ΔG, as any sentencing range $(\rho^{+}(\tilde{\tau})-\rho^{-}(\tilde{\tau}))$ is positive by definition. To determine circumstance under which the $\Delta G<0$, we plug values of Gs into a ΔG:

$$\Delta G<0\Leftrightarrow G(x,t+\epsilon;\tilde{\tau}_2,\bar{q}_2)<G(x,t;\tilde{\tau}_1,\bar{q}_1)\Leftrightarrow$$
$$\frac{\int_{\tau_1}^{\tau_2}\int_{x_{min}}^{q^{-1}(\bar{q},l+\epsilon)}f_{X,T}(s,l)\mathrm{d}s\mathrm{d}l}{\int_{\tau_1}^{\tau_2}\int_{x_{min}}^{x_{max}}f_{X,T}(s,l)\mathrm{d}s\mathrm{d}l}<\frac{\int_{\tau_0}^{\tau_1}\int_{x_{min}}^{q^{-1}(\bar{q},l)}f_{X,T}(s,l)\mathrm{d}s\mathrm{d}l}{\int_{\tau_0}^{\tau_1}\int_{x_{min}}^{x_{max}}f_{X,T}(s,l)\mathrm{d}s\mathrm{d}l} \tag{1}$$

Figure 1 shows the intuition behind the inequality. To satisfy the condition, we need the share of offenses (x,t) in the bottom segment on the left to be higher than the share of offenses on the bottom segment on the right. In other words, the condition would

[9] In fact, the *severity effect* is almost always positive.

Figure 1: Negative Reference Effect



Notes: Figure represents the intuition behind the equation 1. For the severity effect to be negative, it is necessary that the share of cases below the red line on the left from the τ_1 is smaller that the share of cases on the right below the red line. Note that the negative slope of the red line is given by assumption 1: monotonicity of seriousness function.

be violated if among the cases with higher classifying variable were systematically more cases with lower circumstances x, i.e. correlation $corr(X, T)$ was negative enough, which seems unlikely.

2.3.1 Implications

The framework introduced allows us to discuss how the structure of the thresholds shape sentencing more generally. We limit the discussion to one/two particular implication(s).

In some countries, such as in France, the sentencing ranges are organized so the lower limit of the sentencing range is the same across more (all) subsections of the offense and only the upper limit increases for subsections for higher values of classifying variables. Our framework helps us to understand how the *severity* and the *reference* effects will differ compared to a system of thresholds in which both the lower and the upper thresholds increases for subsections for higher values of classifying variables.

Formally, suppose two systems denoted as α and β, so that the sets of thresholds are identical ($\mathcal{T}^{\alpha} = \mathcal{T}^{\beta} = \tau^{[0]}, \tau^{[1]}, \tau^{[2]}, \ldots$) and also the upper limits of sentencing ranges are the same, $\rho^{+}_{\alpha}(\tau) = \rho^{+}_{\beta}(\tau)\ \forall\ \tau$. The two systems differ in the lower limit of the sentencing ranges. In particular, in the α system, the lower limits of sentencing ranges remain the

same for all subsections $\rho_{\alpha}^{-}(\tau^0) = \rho_{\alpha}^{-}(\tau^1) = \rho_{\alpha}^{-}(\tau^2)$, while the β system, the sentencing range is increasing $\rho_{\beta}^{-}(\tau^0) < \rho_{\beta}^{-}(\tau^1) < \rho_{\beta}^{-}(\tau^2)$.

Our framework suggests that the severity effect is weaker in the α system and it is thus more likely that the case with marginally higher classifying variable will be sentenced to more lenient sentence than in the system β. This comparative statics assumes that the $G(x, t; \tau, \bar{q})$ is the same in both systems.

3 Experiment with Prosecutors

The aim of our experiment is twofold. First, the design allows us to quantify the consequences of thresholds in a controlled environment with professional sentencers and determine Δs. The theoretical model reveals that under reasonable circumstances, the overall effect is a sum of two opposite mechanisms and the final sign and magnitude of the effect of the threshold is an empirical question. Second, we design an experimental treatment to test existence of the severity and reference mechanisms.

3.1 Background

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

Prosecution's role in sentencing consists in recommendation of sanctions and in wide possibilities to appeal against a sentence. While until 2019 practices of prosecutors diverged as to whether they would recommend a sanction already in the indictment or only in the closing argument, from 2020 they are obliged to recommend a specific sanction together with the indictment. Prosecutors can file an appeal arguing either an error of law or fact was made; no specific level of an error (i.e. substantial or palpable) is required. Judges are highly incentivized to appease both the defendant and the prosecutor so that they waive their right to an appeal: In such case, judges can issue a simplified judgment and do not have to provide a detailed written reasoning, saving them several hours of work. Prosecutors are well aware of these powers as they have employed them in 2016 to achieve a high increase in fine imposition rate (Drápal & Dušek, 2020).

Regarding mentality Czech Prosecutors are not law-and-order punitive players, illustrated by Prosecutor General's Office suggesting several ways of reducing high prison population in the Czech Republic, such as via lowering of sentencing ranges. Regarding organizational structure, while Prosecutor General can issue binding orders in general matters (such as regarding recommendation of a specific sentence), it generally cannot interfere with individual cases. This is due also to a fourth-level hierarchical structure of prosecution copying the judicial structure: District courts dealing with vast majority (98 %) of cases, regional courts trying the most serious cases and serving as appeal courts against decisions of district courts, high courts serving as appeal courts against

decisions of regional courts and Supreme Court deciding on appeals in legal matters. The head of a higher prosecution office can directly influence only the cases dealt by a prosecution office of one lower level, making regional prosecution offices the most important for influencing every-day practice. Head of each prosecution service then influences sentence recommendations as they approve all indictments and sentence recommendations. The Prosecutor General thus has only limited role in influencing sentencing practices of first-level prosecutors.

Prosecutors are not elected via a popular vote, but they are appointed by a Minister of Justice following a proposal by the Supreme Prosecutor after serving at least 3 years as assistants to a prosecutor, a judge or an attorney and after passing a professional exam. They have tenure and they can be removed only via disciplinary proceedings if they commit blatant mistakes. While above we focused on their role in sentencing, prosecution focuses primarily on overseeing investigations, filing indictments and participating at court hearings. Anecdotal evidence offered by several prosecutors and judges suggests that more attention is paid to the decision on guilt than to the one on sentence.

3.2 Experimental Design

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

Theorem 1 implies that to test for the reference and severity effect requires that the other mechanism is muted. To test for the severity effect, the reference effect must be suppressed and *vice versa*. To attain such comparison, we vary the sentencing ranges and offense compositions and introduce hypothetical situations.

In particular, to test for the reference effect, one needs to compare scenarios, in which two identical cases are sentenced according to subsections with the same sentencing ranges, but with a different structure of cases. Formally, $\Delta\rho^- = 0$ and $\Delta\rho^+ = 0$, while $\Delta G \neq 0$. We generate the variation by adjusting the thresholds and thus effectively adding more severe criminal cases into the comparable structure of cases.

Similarly, to test for severity effect, we compare two scenarios for which the relative position among other cases remain the same ($\Delta G = 0$), but the upper boundaries of the sentencing ranges increase. To minimize inconsistency in the criminal code in this

hypothetical situation, we restrict ourselves to only small changes in sentencing range. Such changes increase the credibility of hypothetical scenario, as the hypothetical scenario resembles the current situation known to prosecutors, but it also diminishes the predicted effect.

3.3 Implementation of Experimental Design

To approach prosecutors we partnered with the Prosecutor General's Office of the Czech Republic. The invitation to participate in the survey was sent by a deputy to the Prosecutor General. Initially 1257 prosecutors were invited to participate and 206 prosecutors responded to the first vignette and 194 to both of them. Each of the participant received a unique link, enabling us to eliminate duplicate answers from the same prosecutors. Additionally, the unique link allows us to match the experimental data with anonymized administrative data containing basic information about prosecutors such as their gender, age, time at bench, position in organization of prosecutors, alma mater and whether they were members of the communist party up to the 1989 Velvet Revolution.

The sample of participants is not a representative sample of the population of prosecutors. Compared with the population of prosecutors, our sample contains a higher proportion of female respondents, those who work at the Prosecutor General's Office, and participant are, on average, younger. It is likely that the overrepresentation of the Prosecutor General's Office's prosecutors is caused by higher willingness to comply with the request of their direct superior. The descriptive statistics are presented in Appendix in Table 6.

Each participant in the experiment was asked to analyse and recommend a length of incarceration in two hypothetical criminal cases: (i) drug possession and (ii) theft. The cases were presented in this particular order to everyone. Each case was described on 1 page and the wording was consulted with several practising judges, so no relevant informant is omitted. We also provided the respondents with the relevant offense section of the criminal code (with multiple subsections) together with the relevant section of the criminal code or excerpt from jurisprudence establishing the thresholds. They were warned on the introductory screen that the provisions my not be in line with those in force; this comment was emphasized in italic. Additionally, we asked the respondents to indicate according to which subsection they sentence the offender. Since the classification of the subsection is based on quantifiable variables that were provided in the case, there is objectively right answer. Failure to identify the correct subsections can be a sing of inattentive of careless responses.

Participants were presented with an order from the head of their prosecution office to impose non-suspended prison sentence, while its length was left entirely up to them.

The cases were prepared in such a way that the imposition of a non-suspended prison sentence would be the most likely choice for most prosecutors due to the extensive criminal history of defendants. To select the length of imprisonment, participants were supposed to choose both number of years and months of imprisonment from a drop-down menu. Their answers were not limited to the relevant length of incarceration according to the criminal code, so it was technically possible to recommend longer or shorter sentenced than given by the criminal code.

Each criminal case was presented in four different scenarios (i.e. there are four treatment arms in both cases) that differ in parameters of the cases and generates the intended experimental variation. Upon entering the platform, participants were randomized into one of the treatment arms in the first case (drug possession). The randomization into a treatment arm in the second case (theft) was independent from the first randomization. Note that since not everyone who started with the survey also finished it, the randomization based on entry does not, necessary, lead to numerically identical number of observations in each scenarios. We thus implement a between subjects design and will explore the variation in length of incarceration recommended in different treatment arms by different respondents.

3.4 Drug Possession

3.4.1 Vignette description

In the drug possession case, participants were asked to recommend a prison sentence for an offender, who was selling methamphetamine in front of a dance club in a town in Northern Bohemia. It was a one-time event following the offender being fired from a job. He was repeatedly sentenced for selling marijuana over the last 10 years and for small thefts. We have manipulated the amount of methamphetamine (its pure substance, henceforth only methamphetamine) found on him, the sentencing range applicable and the thresholds influencing composition of cases within subsections. The full text of the vignette together with the text of the section and other information provided to the participants is presented both in English and in Czech in Appendix. The four scenarios that were applied are summarized in Table 1. Scenario A and scenario B are based on the existing legislation and differ marginally in the amount of the drug possessed. The difference between the length of incarceration recommended in scenario A and B represents the overall effect causes by the threshold. Plugging the parameters into theorem 1 yields $s(B) - s(A) = 1 - 4G(A) + 8G(B)$.[10]

[10] We employed simplified notation, in which $G(A)$ means function G for parameters applied in scenario A.

Table 1: Scenarios in Drug Possession Case

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

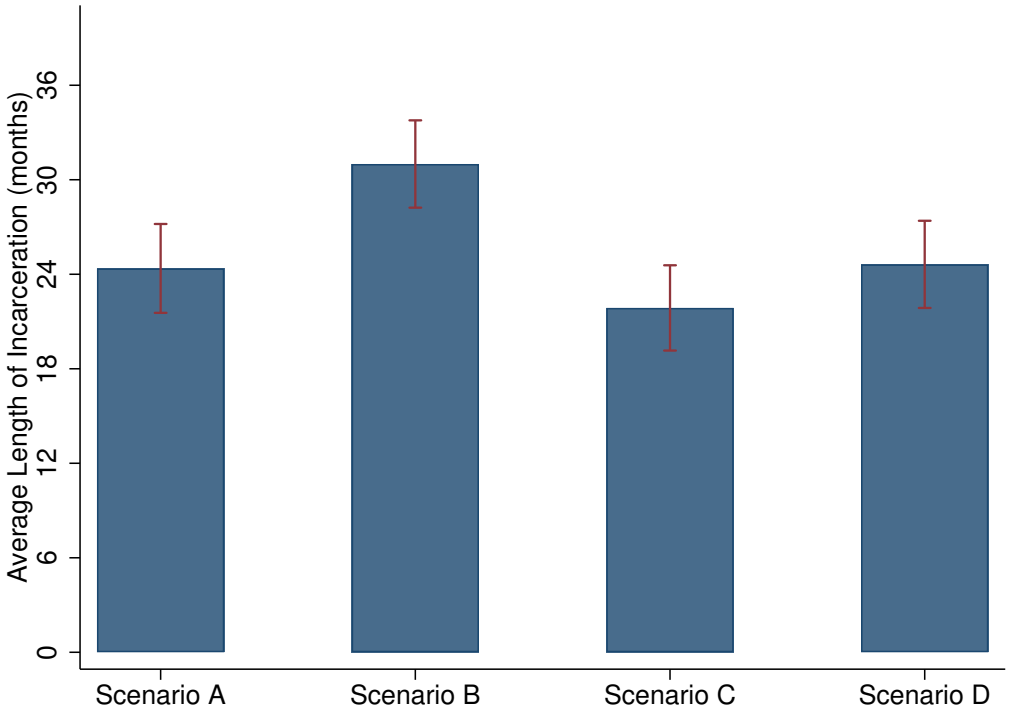
Scenarios C and D introduce alternative legal framework. In particular, in scenario C we move the upper threshold of classifying variable from 150 g to 300 g of methamphetamine and thus we add relatively more severe cases in the relevant subsection. It follows that a comparison between A and C scenarios represents a reference mechanism. Formally, $s(C) - s(A) = 4\,(G(C) - G(A))$. Note that if $G(A) > G(C)$, the effect is negative. In scenario D, we move the upper limit of the sentencing range and thus evoke the severity mechanism. Note, however, that the magnitude of the reference mechanism estimated as a comparison between scenario A and D is only a fraction of the whole severity effect that works in the overall effect. In fact, we estimate $s(D) - s(A) = 3G(A)$.

3.4.2 Results

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

Table 2 shows results from formal test of three hypotheses; panel I tests the overall (main) threshold effect, panel II tests the reference effect, and panel III tests the severity effect. Each panel shows results from three models. Model 1 represents a simple univariate OLS regression in which the treatment effect is captured by a dummy that equals to 1 if the observation comes to the treated scenario and 0 otherwise. The latter conveniently corresponds to scenario A in all three panels of the table. See equation 2. Model B extends the univariate OLS by controlling for additional characteristics of prosecutors (age, age^2, gender, position in the system of prosecutors). Finally, model C estimates an univariate OLS that includes responses also from participants who did not identify the correct subsection.

Figure 2: Length of Incarceration Suggested by Prosecutors - Drug Possession



Notes: The graph shows the average length of incarceration suggested by prosecutors in four scenarios of the drug possession case. In scenario A, the offender was caught with 147.8 grams of methamphetamine, while the threshold is 150 g and sentencing ranges from 1 to 5 years of incarceration. The average sentence in scenario A is 24.4 months. In scenario B, the offender was caught with 151.8 grams of methamphetamine, while the threshold is 150 g and sentencing range from 2 to 10 years of incarceration. The average sentence in scenario B is 31 months. In scenario C, the offender was caught with 147.8 grams of methamphetamine, while the threshold is 300 g and the sentencing range from 1 to 5 years of incarceration. The average sentence in scenario B is 21.9 months. In scenario D, the offender was caught with 147.8 grams of methamphetamine, while the threshold is 150 g and sentencing ranges from 1 to 8 years of incarceration. The average sentence in scenario B is 24.6 months. The lines represent 95 % confidence intervals.

$$\text{Sentence} = \alpha + \beta \text{Effect} + \epsilon \tag{2}$$

Panel I of Table 2 provides convincing evidence that the threshold increases the length of incarceration recommended by around 6 months, which represents approximately 25 % increase of the length. Employing different models the point estimates of the treatment effect range from 5.2 to 6.6 months. The results thus provide robust evidence that two almost identical cases (the difference is 3 grams of methamphetamine (2.7 %)) were recommended to a significantly different length of incarceration.

Panel II of the Table 2 shows suggestive evidence of the reference mechanisms. All three specifications provide similar negative point estimates, suggesting that by increasing the upper threshold for classifying variable from 150g to 300g the sentence decreases by 2.5 - 3 months. The effect represents approximately a 10 % increase in the length of incarceration. However, since the point estimates are not statistically significant, one needs to be causation with the interpretation.

Finally, three specifications in panel III test for the severity effect and suggest the null results. To understand why the severity effect is rather negligible, it is important to realize that it does not test the full severity effect, but rather existence of the mechanism. The average sentence in scenario A is located around a fourth of the possible interval (a year above the minimum of the sentencing range and 3 years below the maximum of the sentencing range) so increasing the upper limit of the sentencing range further has only a limited impact.[11] Importantly, since the overall threshold effect is sizeable and positive, our framework implies that the full severity effect must be sizeable and positive.

Studying the average values masks a substantial heterogeneity in prosecutors behavior. Panel A and B of figure 6 in Appendix show histograms of all decisions in scenarios A and B, respectively. Most of the suggested length in scenario A are concentrated in the lower part of the sentencing range of 1 to 5 years and 9 of the prosecutors even decided to suggest the minimum length. In scenario B, the pattern is similar, as none of the prosecutor suggested a sentenced above the middle length of the sentencing range (72 months) and 23 (38 %) of the prosecutors suggested the minimum of 24 months. Interestingly, two prosecutors suggested a sentence below the minimum.[12]

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

[12] Furthermore, the figure shows that prosecutors tend to round the length to years which is in line with previous research Dhami et al. (2020)

Table 2: Marginally Higher Amount of Drugs Possessed Increases Sentence Significantly

Panel I: Effect of Threshold

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

Panel II: Reference Effect

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

Panel III: Severity Effect

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

Robust standard errors in parentheses

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

Panel I tests the overall threshold effect (s(B) - s(A)); panel II tests the reference effect (s(C) - s(A)); and panel III tests the severity effect (s(D) - s(A)).

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

3.5 Theft

3.5.1 Vignette description

In the theft vignette the prosecutors were asked to recommend a prison sentence for an offender visiting his parents to help them with updating their computer. He was left alone in their apartment, took advantage of his parents being logged-in in their internet banking account and of their cellphone being left at home and transferred all their money to his own account. He gambled away all of it in the following three days. He was previously sentenced for fraud, embezzlement and he was released from a two-year prison sentence two months prior committing the offense in the vignette. The entire vignette with text of the sections of the criminal code are presented in Appendix. Similarly to the previous case we randomized participants into four treatment arms (scenarios), as captured by table 3. Unlike in the previous case, we did not include scenarios that vary the existing legislation. Instead, we study overall threshold effects around two sequential thresholds, labeled as the 50k and the 500k thresholds.

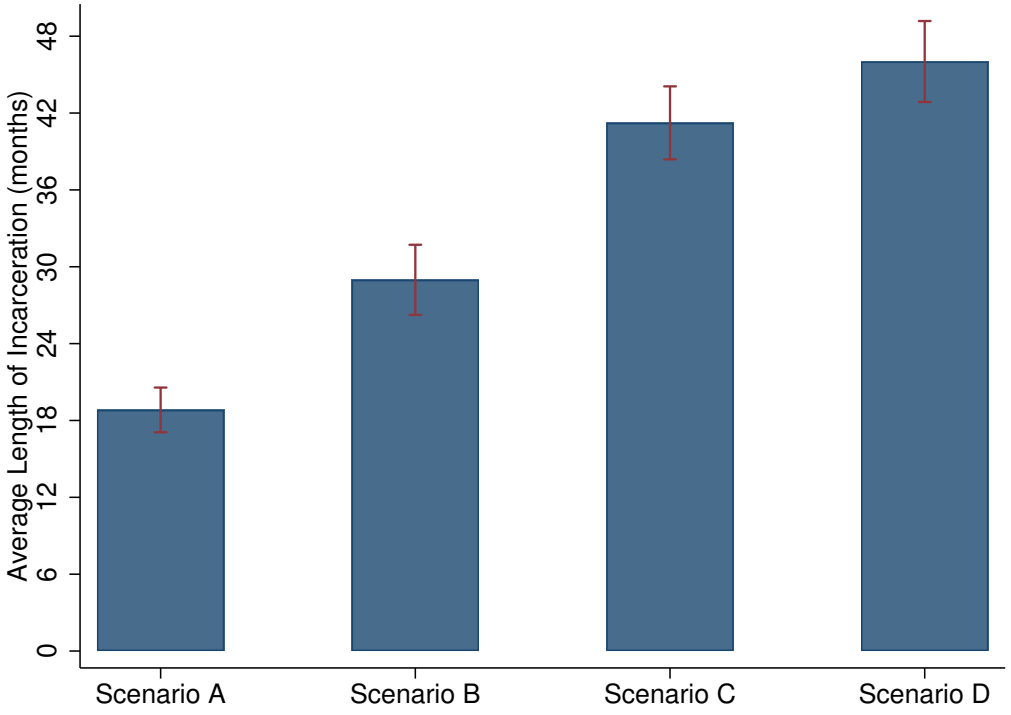
Table 3: Scenarios: Theft

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

We estimate three causal effects regarding the theft case. First, a comparison of the average length of incarceration recommended in scenarios A and B yields the effect of the 50k threshold. The difference corresponds to $s(B)-s(A) = 1-2G(A)+4G(B)$. Second, a comparison between scenarios C and D estimates the change of the length of incarceration caused by the 500k threshold. The effect corresponds to $s(D)-s(C) = 1+6G(D)-4G(C)$. Third, we investigate what is the effect of increasing the classifying variable dramatically. Scenario B and C differ only in the size of the damaged caused, so the comparison between the average length recommended estimates the causal effect of dramatic increase in damaged caused. Formally, the effect estimated is $s(C) - s(B) = 4\,(G(C) - G(B))$. Note that since we assume $q(x,t)$ is increasing in t, the effect is predicted to be positive.

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

Figure 3: Length of Incarceration Suggested by Prosecutors - Theft



Notes: The graph shows the average length of incarceration by different scenario. In scenario A, the offender caused a damage of CZK 48,283, while the relevant threshold is CZK 50,000 and the sentencing range from 0 to 2 years. The average sentence in scenario A is 18.8 months. In scenario B, the offender caused a damage of CZK 51,283, while the relevant threshold is CZK 50,000 and the sentencing range from 1 to 5 years. The average sentence in scenario B is 29 months. In scenario C, the offender caused a damage of CZK 487,092, while the relevant threshold is CZK 500,000 and the sentencing range from 1 to 5 years. The average sentence in scenario C is 41.2 months. In scenario D, the offender caused a damage of CZK 508,213, while the relevant threshold is CZK 500,000 and the sentencing range from 2 to 8 years. The average sentence in scenario D is 46 months.

of the magnitude of the effect caused by the 50k threshold in terms of the magnitude of the classifying variable.

3.5.2 Results

Figure 3 shows the average length of incarceration recommended by prosecutors in different scenarios. The higher the damage caused the longer the recommended sentence. In scenario A, the average length of incarceration is slightly above 18 months. In scenario B, the average sentence is by 10 months higher. In the remaining scenarios C and D the average sentences are approximately 41 and 46 months, respectively.

Table 4 shows three panels, each devoted to one of the three questions: the effect of 50k threshold, the effect of 500k threshold, and the effect of size of the classifying variable. Each of the effect is estimated using the same three specifications applied for the drug

possession case. Panel I presents robust evidence that the 50k threshold increases the average incarceration by about 10 months, which represents more than 50 % increase compared to the sentence for cases just below the 50k threshold. Once we control for gender, age, and position in the hierarchy prosecutors the point estimates are even larger.

Panel II estimates the effect of the 500k threshold. Both model 1 and model 2 suggest that the 500k threshold increases the average length of incarceration by 4.5 months (app. 10 %). While the effect of the 500k threshold is smaller than the one caused by the 50k threshold, it provides additional evidence that existence of the effect is rather general and limits the external validity concerns. Finally, panel III presents evidence that increasing the damage almost 10 times increases the sentence by roughly a year.

The effect of size of the damage caused allows us to understand the enormous effect the thresholds have. Increasing a damage by 450,000 CZK (by 830 %) corresponds to additional 12 months of incarceration. This contrasts to 10 months increase of incarceration for a theft that causes higher damage by 3,000 CZK estimated in panel A. Alternatively, we can compare the 50k threshold effect to the effect causes by increasing the damage caused approximately 10 times. Formally, we run the following regression

$$\text{Sentence} = \beta_1 \text{Scenario A} + \beta_2 \text{Scenario B} + \beta_3 \text{Scenario C} + \varepsilon \tag{3}$$

and using the Wald test, we test

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

Not rejecting the null hypothesis suggests that the 50k threshold has the same effect as increasing the damage caused from by 450,000 CZK (by 830 %). Despite the probable sensitivity of the results on particular specifications and circumstances of cases, we provide robust evidence that the effect of thresholds on the shape of sentencing decisions is of a great importance.

Table 4: Marginally Higher Amount of Damage Caused Increases Sentence Significantly

Panel I: 50k Threshold

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

Panel II: 500k Threshold

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

Panel III: Size of Damage

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

Robust standard errors in parentheses.

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

Panel I tests the 50k threshold effect (s(B) - s(A)), panel II tests the 500k threshold effect (s(D) - s(C)), and panel III tests the effect of size of damage (s(C) - s(B)).

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

4 Measure of Justice and Injustice

4.1 Relatively Just Decision and Its Measure

To provide results beyond the average recommended sentence, we introduce an alternative measure that quantifies probability of just decision. We first define *just* decision and then propose an empirical measure. To avoid a controversial normative stance on what decisions are *just* and what not, we introduce an ordinal, relative definition.Using the notation introduced earlier, consider a case (x, t) and corresponding sentencing decision decision s_1. Given this sentencing decision s_1, we ask whether a different sentencing decision s_2 in a case of $(x + \delta_x, t + \delta_t)$, with $\delta_x \geq 0$ and $\delta_t \geq 0$ is *just* or not. We say that a sentencing decision s_2 is *just* if the corresponding sentence is: (i) at least as harsh as s_1; and (ii) the punishment is not unreasonably harsher.[13]

Definition 3 (Just Decision)**.** Given a decision s_1 of an offense (x, t) and a tolerance parameter $\eta \geq 1$, we say that a decision s_2 of an offense $(x + \delta_x, t + \delta_t)$, where $\delta_x \geq 0$ and $\delta_t \geq 0$ is *just* if the two following conditions are satisfied:

$$s_1 \leq s_2 \tag{D 3.1}$$

$$s_2 \leq s_1\eta. \tag{D 3.2}$$

Importantly, the definition says that a decision s_2 can be called *just* only if there is a tolerance parameter η and a decision s_1. In other words, the justice of any sentencing decision is viewed relative to other sentencing decision and a tolerance parameter. Without these, the definition is meaningless.

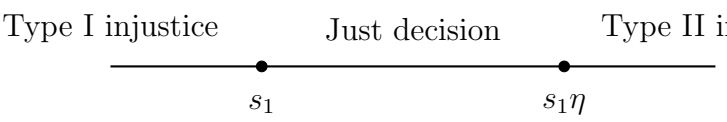
The definition of *just* decision implies that sentencing decisions that are not considered *just*, can be of two types. Depending on which condition is not satisfied, we distinguish two types of injustice. If condition D 3.1 is not satisfied and the more severe case is sentenced to less harsh punishment, then we talk about *type I injustice*, whereas if condition D 3.2 is not satisfied and the sentence s_2 is too harsh, we talk about *type II injustice*. Figure 4 represents the idea.

Having defined *just decision* and types of injustice, we next define empirical measures. To fit our experiment, we will consider the following situation. There are N_i sentencing decisions in a case (x, t) and call each of the decisions s_i. There are also N_j decisions in cases $(x + \delta_x, t + \delta_t)$ denoted as s_j. Suppose that sentencing decisions are made with a noise. In particular, instead of a sentence s, the realized decision is $\hat{s} = s + \xi$.[14]

[13]Note that this approach follows Aristotle's principle to decide alike cases alike and different differently.

[14]There are several different interpretation of the noise. For example, it can represent a heterogeneous preferences or tastes in sentencing decisions among sentencers.

Figure 4: Just Decision



Notes: Given a parameter of tolerance η and a sentencing decision s_1, then depending where the sentencing decision s_2 falls we talk about a *just* decision, *type I injustice* or *type II injustice*.

We introduce three empirical measures. The measure of justice $\mathcal{M}^J(\eta)$ compares each sentencing decision s_i to each sentencing decision s_j and quantify frequency of $\frac{\hat{s}_j}{\hat{s}_i}$ being between 1 and η. Similarly, one can define a measure of *type I injustice* $\mathcal{M}^I$ and *type II injustice* $\mathcal{M}^{II}(\eta)$. Note that *type I injustice* does not depend on the tolerance parameter η. Intuitively, our measures quantify the probability that for a pair randomly observed sentences s_i and s_j, the latter will be less severe, reasonably harsher or unreasonably harsher compared to the former. The following definition introduces all three measures.

Definition 4 (Measures of Justice and Injustice)**.** Suppose that there are N_i decisions in a case (x,t) denoted s_i and N_j decisions in cases $(x+\delta_x, t+\delta_t)$ denoted as s_j, in which $\delta_x \geq 0, \delta_t \geq 0$. Suppose further that the observed decisions are made with a noise in a form $\hat{s_i} = s_i + \xi_i$ and $\hat{s_j} = s_j + \xi_j$. Then for a given parameter η, the measure of justice $\mathcal{M}^J(\eta)$, measure of type I injustice $\mathcal{M}^I(\eta)$, and measure of type II injustice $\mathcal{M}^{II}(\eta)$ are

$$\mathcal{M}^J(\eta) = \frac{\sum_{i\in I}\sum_{j\in J}\mathbf{1}\left[1 \leq \frac{\hat{s}_j}{\hat{s}_i} \leq \eta\right]}{N_J N_I} \qquad \text{(Measure of Justice)}$$

$$\mathcal{M}^I = \frac{\sum_{i\in I}\sum_{j\in J}\mathbf{1}\left[\frac{\hat{s}_j}{\hat{s}_i} < 1\right]}{N_J N_I} \qquad \text{(Measure of Type I Injustice)}$$

$$\mathcal{M}^{II}(\eta) = \frac{\sum_{i\in I}\sum_{j\in J}\mathbf{1}\left[\eta < \frac{\hat{s}_j}{\hat{s}_i}\right]}{N_J N_I}. \qquad \text{(Measure of Type II Injustice)}$$

Prior we quantify the measures for the experimental cases, we discuss a few properties. First, the sum of the measures equals to 1 for any tolerance parameter η. Second, the higher the tolerance parameter η is, the higher the *measure of justice* is. Intuitively, as we increase the tolerance parameters even harsher sentences are considered just (or tolerated). Naturally, by increasing the tolerance parameter, the measure of *type II injustice* decreases. Third, the measure of *type I injustice* is bounded between $(0, \frac{1}{2})$. As both $\delta_x \to 0$ and $\delta_t \to 0$ i.e. cases (x,t) and $(x+\delta_x, t+\delta_t)$ are alike, the measure of *type I injustice* converges to $\frac{1}{2}$. On the contrary, the more the cases differ, the measure of *type I injustice* converges to 0. This property can be used to empirically identify perceived

similarity among cases by sentencers.

Proposition 1 (Properties of Measures)**.** Consider measures introduced in definition 4 and suppose that ξ_i and ξ_j are distributed according to the same cumulative distribution function Ξ. Suppose further η and η': $\eta < \eta'$. The measures have the following properties

$$\mathcal{M}^J(\eta) + \mathcal{M}^I + \mathcal{M}^{II}(\eta) = 1 \quad \forall\eta \tag{P 1.1}$$

$$\mathcal{M}^J(\eta) < \mathcal{M}^J(\eta') \tag{P 1.2}$$

$$\mathcal{M}^I \in (0, \frac{1}{2}) \tag{P 1.3}$$

4.2 Measure of Justice in Experiment

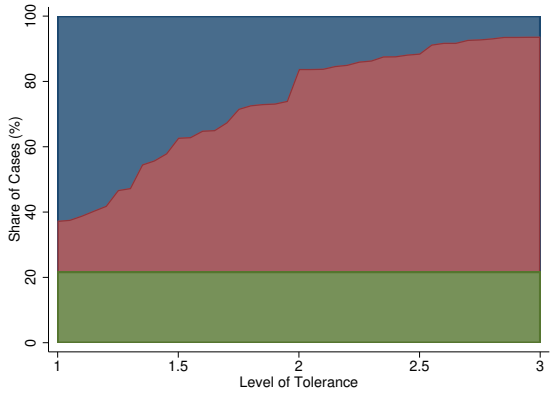
Using the definition 4, we calculate measures for four comparisons: (i) scenario A and B in the drug possession case; (ii) scenario A and B in the theft case (50k threshold); (iii) scenario C and D in the theft case (500k threshold); and (iv) scenario B a C in the theft case (Size of Damage). Table 5 reports four statistics for each of the comparisons: *type I injustice* measure, *measure of justice* for tolerance parameter 1.5 and 2, respectively. Finally, an inverse function of *measure of justice* evaluated at 0.5 that says how tolerant one has to be (i.e. what the η must be) to reach 50 % probability that from a randomly observed pair s_i and s_j, the latter is considered a just decision. For example, the first row that compares scenario A (148.8 grams) and scenario B (151.8 grams) of the drug possession case and reveals that to reach 50 % of just decisions, one has to consider 75 % longer prison sentence for 3 grams of methamphetamine as a just decision.

Table 5: Measure of Justice in Experiment

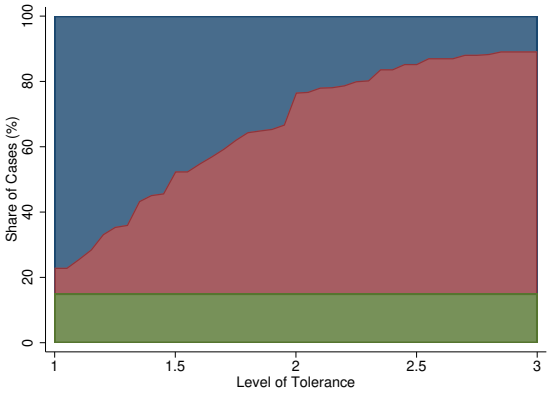
	Type I injustice	$\mathcal{M}^J(1.5)$	$\mathcal{M}^J(2)$	$\left(\mathcal{M}^J\right)^{-1}(0.5)$
Drug Possession	0.21	0.41	0.62	1.75
Theft (50k threshold)	0.15	0.38	0.62	1.85
Theft (500k threshold)	0.33	0.49	0.63	1.55
Size of Damage	0.15	0.41	0.65	1.70

The measure of *type I injustice* implies that the perceived difference between scenario A (48 283 CZK) and scenario B (51 283 CZK) in the theft case is the same as the perceived difference between scenario B (51 283 CZK) and scenario C (487 092 CZK). In fact, the measure of *type I injustice* of these comparisons are numerical identical. This is additional evidence that the threshold has an enormous effect not only on the average sentence recommended, but also on the probability that two cases will be perceived as similar. Note that in the analysis of average recommended sentence we conclude that they are the same. Figure 5 shows measure of justice and injustice for four main comparisons.

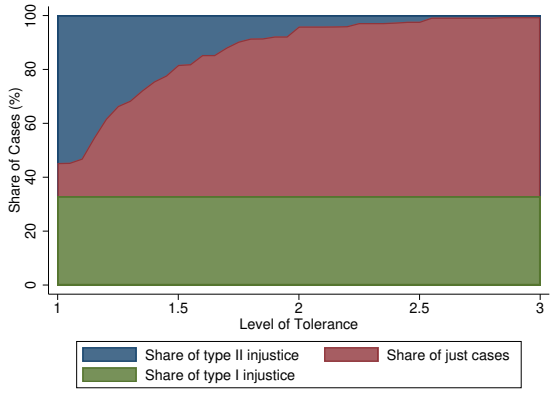
Figure 5: Measures of Justice and Injustice in Experiment



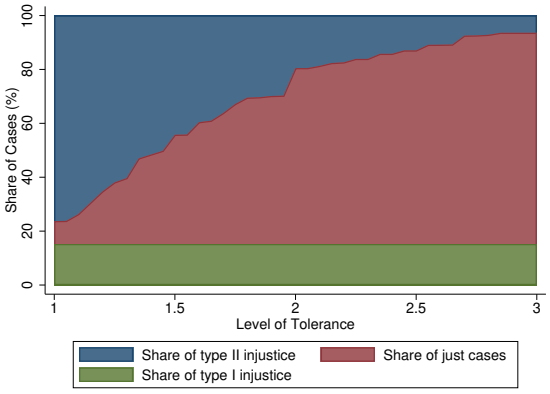
(a) Panel A: Drug Possession



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



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



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

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

5 Discussion

Using a simple model of judges' behavior we model consequences of thresholds dividing offense into categories on sentencing inconsistency. We show that the thresholds give a rise to two opposing effects. Equipped with the model, we design an experiment with Czech prosecutors and empirically demonstrate that the effect of thresholds on sentencing inconsistency is enormous. Falling slightly above the thresholds can lead to a *premium* of up to tens of percentages in length of incarceration. Our results are important in a discussion of how to optimally structure the criminal code.

References

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

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

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

Filipčič, K. (2019). Life imprisonment in Slovenia. *Crimen*, 10(3):225–238.

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

Lappi-Seppälä, T. (2016). Nordic sentencing. *Crime and Justice*, 45(1):17–82.

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

von Hirsch, A. (2017). *Deserved criminal sentences*. Hart Publishing, London.

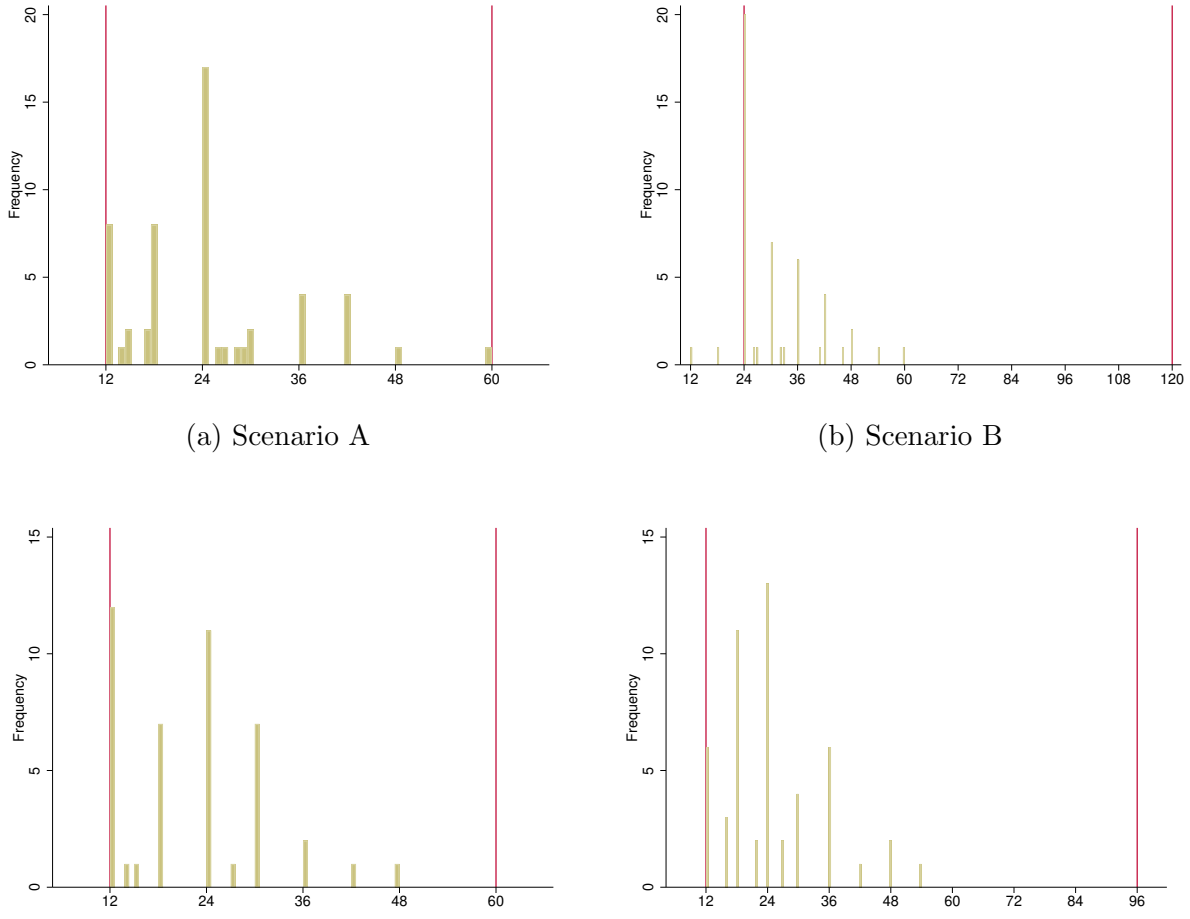
Appendix

Theory

We next show that the equation T 1.1 holds. To simplify notation, we label two cases that are to be compared as A and B and define three differences $\Delta G = G(B) - G(A)$, $\Delta\rho^- = \rho^-(B) - \rho^-(A)$, and $\Delta\rho^+ = \rho^+(B) - \rho^+(A)$.

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

Figure 6: Drug Possession: Length of Recommended Incarceration by Individual Prosecutors



(a) Scenario A

(b) Scenario B

(c) Scenario C

(d) Scenario D

Notes: Each figure represents a histogram of recommended sentences. Red vertical lines mark the upper and low limit of sentencing range. Scenarios correspond to the table 1.

Table 6: Descriptive Statistics

Panel A: Drug Possession

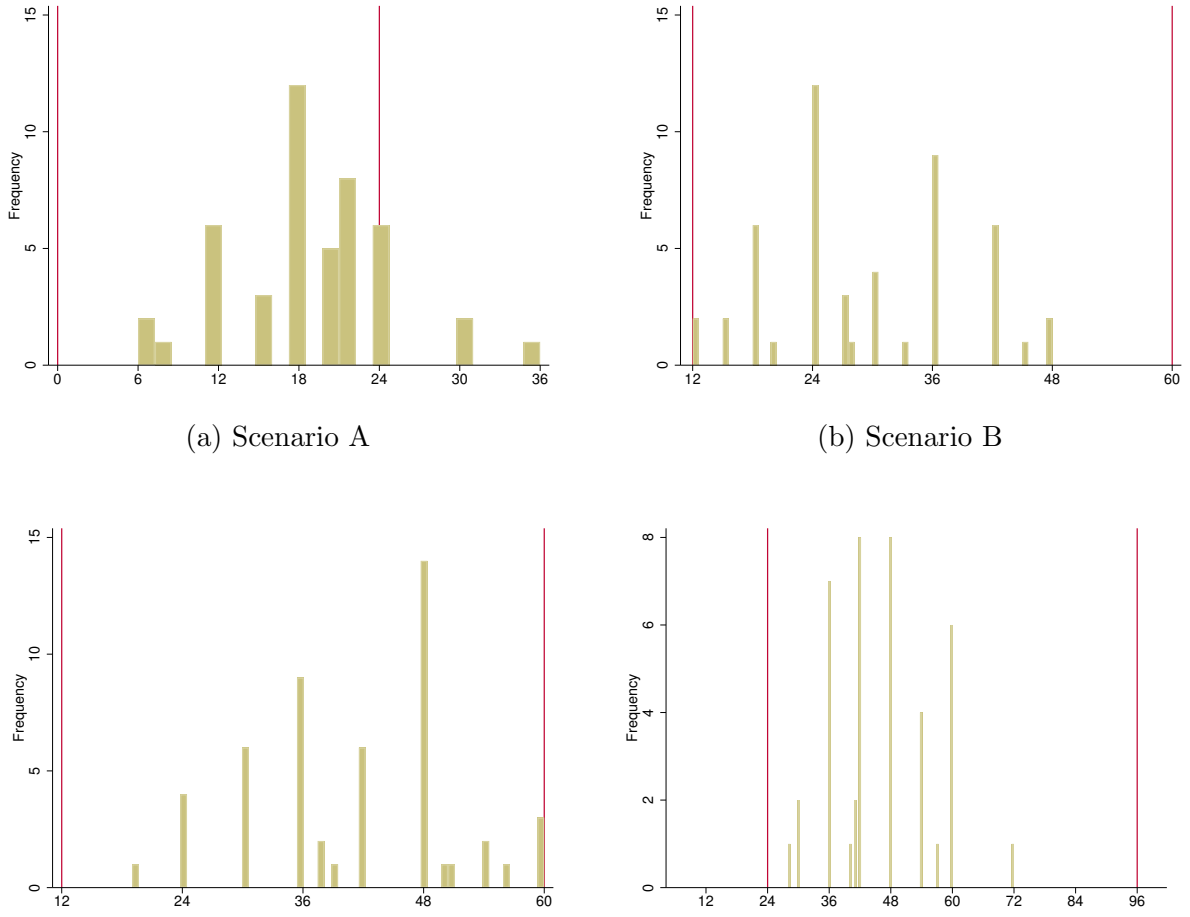
	A	B	C	D	not in experiment	H_0
Number of Observations	58	50	44	54	1049	
Male	41 %	42 %	34 %	52 %	55 %	0.001
Communist Party	10 %	18 %	10 %	15 %	14 %	0.841
Age	46.3	47.8	45.2	47.3	49.1	0.002
Tenure Exam	18.5	19	16.6	20.3	20.4	0.066
Tenure Oath	16.2	16.6	13.8	16.4	17.1	0.071
Alma Mater						
Brno	38 %	38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

Panel B: Theft

	A	B	C	D	not in experiment	H_0
Number of Observations	46	51	51	46		
Male	43 %	27 %	51 %	54 %		
Communist Party	20 %	16 %	6 %	13 %		
Age	48	46.6	47	46.2		
Tenure Exam	19.8	19	18.4	18.5		
Tenure Oath	16.7	18.9	15.9	15.2		
Alma Mater						
Brno	37 %	37 %	37 %	35 %		
Prague	39 %	41 %	37 %	48 %		
Plzen	7 %	14 %	10 %	4 %		
Olomouc	9 %	0 %	8 %	4 %		
NSZ	11 %	14 %	4 %	9 %		
VSZ	4 %	2 %	4 %	9 %		
KSZ	15 %	14 %	20 %	20 %		
OSZ	70 %	71 %	73 %	63 %		

Notes: The H_0 column reports p-value of two-sided t-test under the null that the sample value equals to the population (universe of all active prosecutors) value.

Figure 7: Theft: Length of Recommended Incarceration by Individual Prosecutors



(a) Scenario A

(b) Scenario B

(c) Scenario C

(d) Scenario D

Notes: Each figure represents a histogram of recommended sentences. Red vertical lines mark the upper and low limit of sentencing range. Scenarios correspond to the table 3.

Vignettes

Introduction Screen[15]

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

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

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

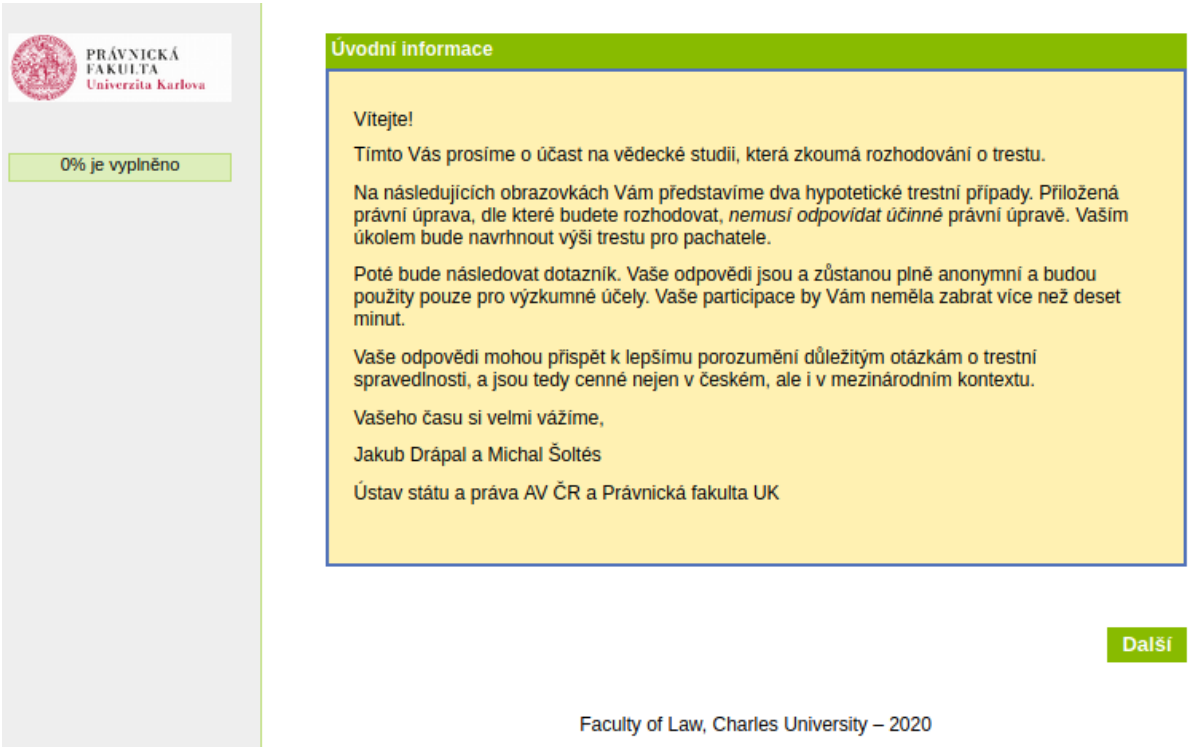
Your answers might contribute to the better understanding of important criminal justice issues that are significant not only for the Czech Republic, but for the international audience as well.

We highly value your time.

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

[15] See figure 8

Figure 8: Vignettes: Introduction Screen



Drug distribution[16]

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

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

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

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

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

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

(1) Whoever produces, imports, exports, transports, offers, provides or sells or otherwise arranges for another or handles for another narcotic or psychotropic substances, products containing narcotic or psychotropic substances, precursors or poisons (meaning 1.5-150 grams of pure substance of methamphetamine according to the jurisprudence of Czech Supreme Court), without an authorisation,

[16] See figure 9

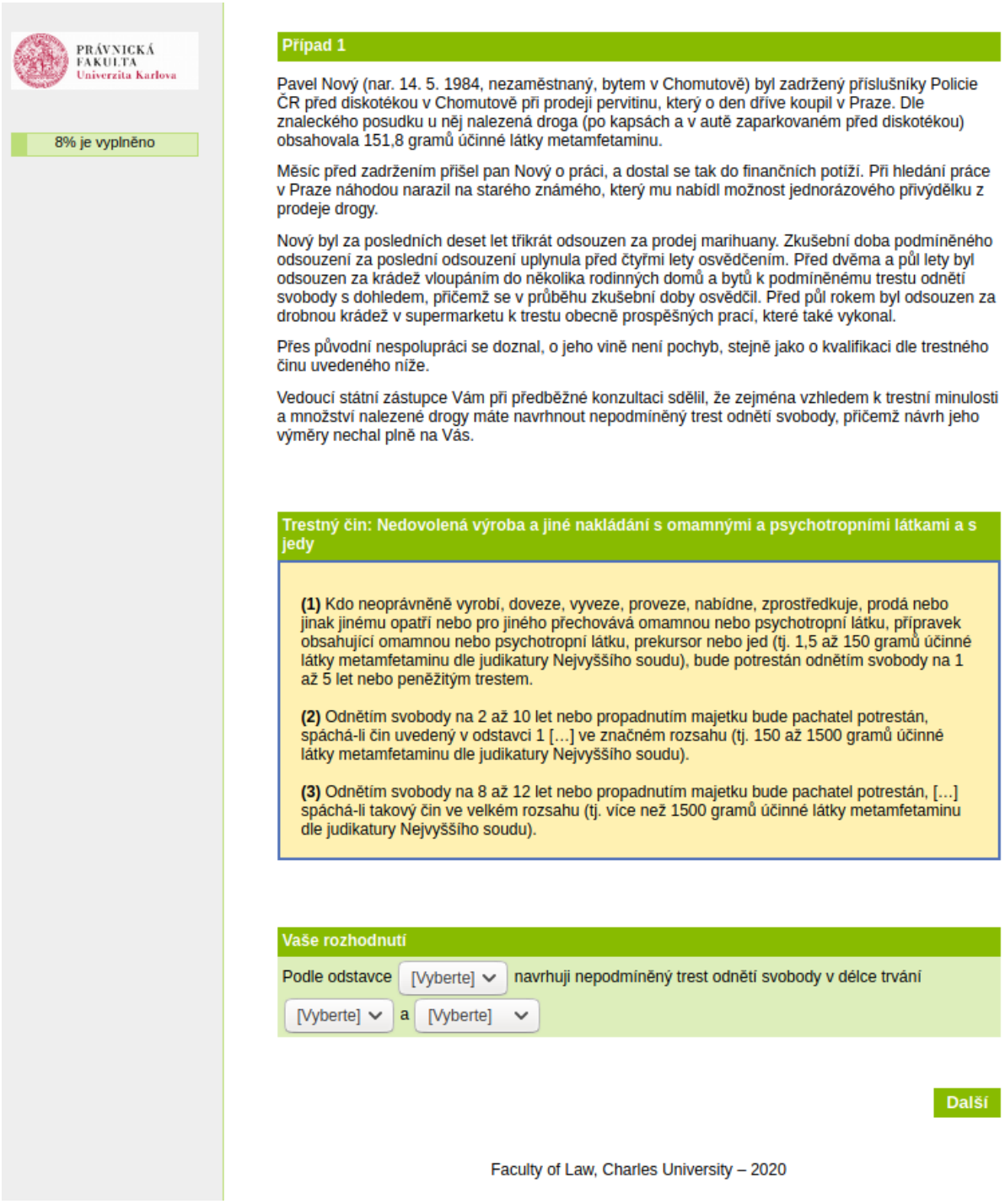
shall be sentenced to imprisonment for one to five years or to a pecuniary penalty.

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

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

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

Figure 9: Vignette 1: Drug Distribution



Vignette 2: Theft[17]

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

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

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

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

Head of Your prosecution office told you during a preliminary discussion that You should recommend non-suspended prison sentence based on offender's criminal record. He left the decision on the length of the non-suspended prison sentence entirely on You.

Offense Section: Theft

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

...

(3) An offender shall be sentenced to imprisonment for one year to five years or to a pecuniary penalty, if he/she causes larger damage (meaning CZK 50,000-

[17] See figure10

500,000) by the act referred to in Sub-section (1).

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

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

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

Figure 10: Vignette 2: Theft