Sentencing Decisions Around Quantity Thresholds: Theory and Experiment*

Jakub Drápal† and Michal Šoltés‡

The most recent version

June, 2021

Abstract

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

*We thank to Libor Dušek, Brian Johnson, Ondřej Kubů, Filip Matějka, Josef Montag, Vladimír Novák, Jiří Pavlík, Ondřej Šťastný, Jan Zápal, and the Prosecutor General's Office and to prosecutors for their participation. 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

†Department of Economics, Faculty of Law Charles University, nám. Curieových 901/7, 116 40 Prague, Czech Republic and Institute of State and Law, Czech Academy of Sciences, Národní 18, 110 00 Prague, Czech Republic (Email: drapalja@prf.cuni.cz)

‡Department of Economics, Faculty of Law Charles University, nám. Curieových 901/7, 116 40 Prague, Czech Republic and 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. (Email: msoltes@cerge-ei.cz)

1 Introduction

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

To render sentencing more consistent and principled, criminal justice systems around the world have introduced various measures (Council of Europe, 1992; Clarkson and Morgan, 1995; Ashworth, 2009). Some of the measures, however, create new sources of disparities. For example, while the U.S. federal sentencing guidelines reduced the level of sentencing disparities at the court level (Anderson et al., 1999), they introduced unwarranted disparities at other levels. Hofer (2019) documents that the guidelines with mandatory minimums lead to higher racial disparities. Furthermore, as prosecutors have gained more discretion, disparities have been further displaced to earlier stages of the criminal process, resulting in large and unjustifiable trial tax (Johnson, 2019) and in charge disparities (Shermer and Johnson, 2010; Tuttle, 2019). While judge consistency may have improved, the overall effect of guidelines is considered less successful. Some even argue that the U.S. federal sentencing guidelines need to be repudiated (Tonry, 2019). The U.S. federal sentencing guidelines serve as a perfect example of how efforts to reduce sentencing disparities may unintentionally cause disparities.

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

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

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

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

Figure 1: Theft Subsections with Sentencing Ranges

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

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

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

To provide empirical evidence, we conduct a Rachlinski-style online experiment with 200 Czech prosecutors. Each participant of the experiment was presented with two vignettes, each describing a different crime case: theft and drug possession. In both cases, we implement several scenarios that vary an amount of classifying variables (amount of methamphetamine possessed and caused damage) around thresholds. We then use the variation to estimate the causal effects of thresholds. Additionally, to test for the existence of the severity and the reference mechanisms, we implement two additional scenarios that introduce an isolated variation in the composition of cases in an offense subsection and sentencing ranges, respectively.

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

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

mechanism provides suggestive evidence that it is negative as predicted.

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

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

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

2 Theory

2.1 Offenses and Thresholds that Divide Them into Subsections

Criminal codes categorize offenses and corresponding sentencing ranges in different ways. In most systems, offenses are divided into subsections with specific sentencing ranges according to certain factors. A finer structure of offenses narrows down sentencers' discretion and provides them with guidance. To divide offenses, criminal codes rely on various factors, some of which are quantifiable such as the amount of damage[3], amount

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

of drugs[4], and number of days when the victim of an assault was not able to work.[5] Quantity thresholds have existed for centuries. For example, theft was divided into subsections with specific sentencing ranges by stolen amount at least since 1803, when the Criminal Code of the Austrian Empire set thresholds for thefts at 25 and 300 ducats.[6] The typical domain of quantity thresholds, however, are drug related offenses (Leader-Elliott, 2012; Sentencing Council, 2011). Despite their popularity, many view them as problematic. In particular, Fleetwood (2011) argues that factors such as the role in drug-trafficking capture offense seriousness better than drug amounts. Furthermore, the amount of drug may be easily manipulated, even by law enforcement officers (Travova, 2019).

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

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

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

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

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

[6]Sections 153 and 159.

range is lower. We call the difference between sentence B and sentence A caused by different composition of cases within subsections as the *reference* effect.

To demonstrate the two mechanisms causing the severity and the reference effects, consider two policy changes that isolate the mechanisms. Suppose that sentencing ranges change, but the composition of the offense subsection remains the same. Then only the severity mechanism is active. Keeping the initial rank of cases ordered by their relative seriousness, sentencers fit the sentences imposed within the new sentencing range. Thus, if a sentencing range for a specific offense subsection is increased (either the lower or the upper limit is increased), sentences are likely to increase as well.

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

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

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

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

offenders who were eligible for it at sentencing (United States Sentencing Commission, 2011, Chapter 8). Altogether, in other systems, prosecutors, offenders, and judges have fewer opportunities and less power to place the case above or below the threshold, to impose sentences below lower sentencing range, and they have less incentive to do so.

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

2.2 Stylized Framework

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

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

Definition 1 (Sentencing Rule)**.** Suppose an offense (x,t). Then the sentence s assigned

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

to this case is determined by the following two-steps *sentencing rule*:

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

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

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

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

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

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

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

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

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

Assumption 2 (An Increase of Sentencing Ranges)

$\forall \tau, \tau' \in \mathcal{T} : \tau < \tau'$, *the following holds*

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

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

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

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

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

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

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

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

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

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

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

[9] See the Appendix for more details.

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

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

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

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

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

2.2.1 Implications

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

Let us denote the two systems as α and β and assume an offense that has two subsections. The thresholds are identical in both systems ($\mathcal{T}^{\alpha}=\mathcal{T}^{\beta}$), Furthermore, the upper limits of sentencing ranges are the same in both systems, $\rho^{+}_{\alpha}(\tau)=\rho^{+}_{\beta}(\tau)\ \forall\tau$. The two systems differ only in the lower limit of the sentencing ranges. In particular, in the α system, the lower limits of sentencing ranges remain the same for both subsections $\rho^{-}_{\alpha}(\tau^0)=\rho^{-}_{\alpha}(\tau^1)$, while in the β system, the sentencing range is increasing. Additionally, we assume that for τ^0 both systems have the same lower limits of the sentencing range. Therefore, $\rho^{-}_{\alpha}(\tau^O)=\rho^{-}_{\alpha}(\tau^1)=\rho^{-}_{\beta}(\tau^0)<\rho^{-}_{\beta}(\tau^1)$. Figure 2 represents the structure of both systems. Note that neither of the systems violates our Assumption 2.

Figure 2: Two different structures of sentencing ranges

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

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

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

3 Experiment with Prosecutors

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

3.1 Background

The Czech criminal justice system is a typical continental European legal system similar to the German one. The Criminal Code and Code of Criminal Procedure establish rules with the higher courts' jurisprudence playing only a quasi-precedential role, as precedents are not binding. The Criminal Code defines both offenses and sentencing ranges for either

the entire offense or for its subsections. As there are no sentencing guidelines, judges have wide discretion as to what type of sanction and of what quantity to impose. Plea bargains are rare, resulting in judges imposing sentences in virtually all criminal cases.

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

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

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

3.2 Experimental Design

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

Theorem 1 implies that testing either for the reference or the severity effects requires that the other mechanism be muted. To test for the severity effect, the reference effect must be suppressed and *vice versa*. To attain such comparison, we introduce a hypothetical legal framework that varies the sentencing ranges and offense compositions. In particular, to test for the reference effect, one needs to compare scenarios in which two identical cases are sentenced according to subsections with the same sentencing ranges, but with a different structure of cases. We generate the variation by adjusting the thresholds and thus effectively adding more severe criminal cases into the subsection. Similarly, to test for the severity effect, we compare two scenarios for which the relative position among other cases remains the same, but the upper limits of the sentencing ranges differ. To minimize inconsistency in the criminal code in these hypothetical situations, we rely only on small changes in sentencing ranges. Such changes increase the credibility of the hypothetical scenario, as it resembles the existing legislation known to prosecutors, but it also suppresses the predicted effect. Tests for the reference and the severity effects are implemented in a drug possession case.[11]

3.3 Implementation of Experimental Design

To approach prosecutors, we partnered with the Prosecutor General's Office of the Czech Republic. The invitation to participate in the survey was sent by a deputy to the Prosecutor General. Of the 1257 prosecutors invited to participate, 206 prosecutors responded to the first vignette and 194 to both of them. Each participant received a unique link, enabling us to eliminate duplicate answers from the same prosecutors. Additionally, the unique link allows us to match the experimental data with anonymized administrative data containing basic information about prosecutors such as their gender, age, time on the bench, position in a prosecutors' organization, alma mater, and whether they were members of the communist party up to the 1989 Velvet Revolution.[12]

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

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

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

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

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

Each criminal case was presented in four different scenarios (i.e. four treatment arms in both cases) that differ in parameters of the cases and generate the intended experimental variation. Upon entering the platform, participants were randomized into one of the treatment arms 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 the survey also finished it; the randomization based on entry does not necessarily lead to a numerically identical number of observations in

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

each scenario. In a between-subjects design we explore the variation in the length of incarceration recommended in different treatment arms.

3.4 Theft

3.4.1 Vignette description

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

The full text of the vignette, along with the text of the section and other information provided to the participants, is presented both in English and in Czech in the Appendix. We randomized participants into four treatment arms (scenarios), as captured by Table 1. In the case of theft, we did not include scenarios that vary the existing legislation. Instead, we study overall threshold effects around two sequential 50k and 500k thresholds.[14]

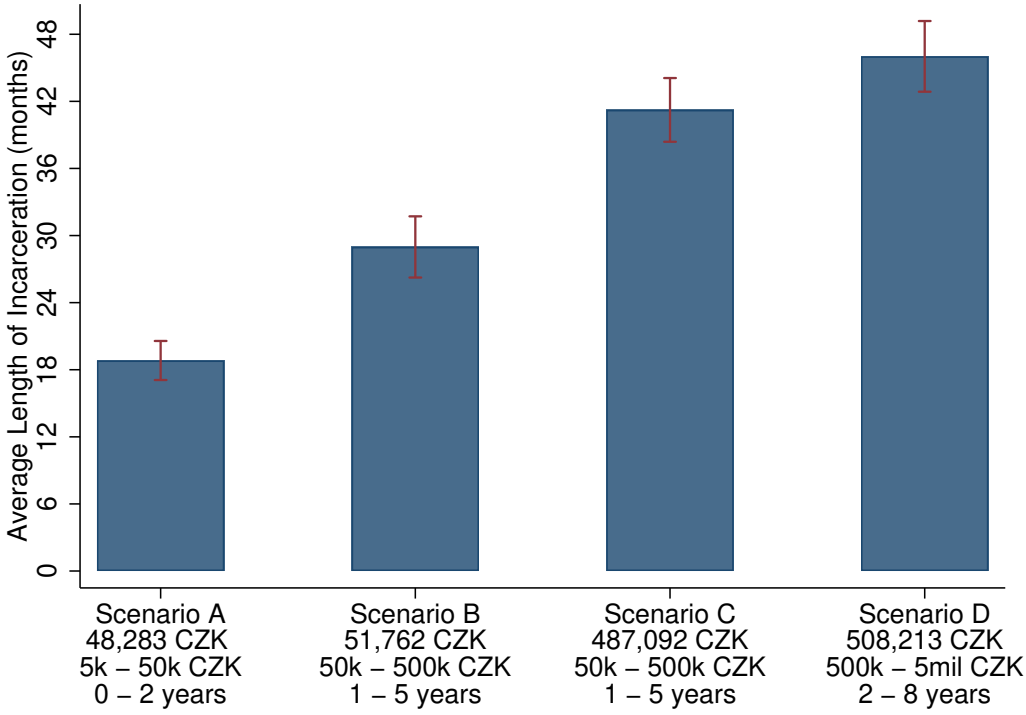
Table 1: Scenarios of Theft Offenses

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

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

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

Figure 3: Theft: Average Recommended Sentence by Prosecutors



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

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

3.4.2 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 higher by 10 months. In the remaining scenarios C and D, the average sentences are approximately 41 and 46 months, respectively.

To test the effects formally, we estimate three models. Model 1 represents a simple univariate OLS regression in which the treatment effect is captured by a dummy that equals to 1 if the observation comes from the treated scenario and 0 otherwise. For the formal regression, see equation Model 1. Model 2 extends the univariate OLS by

controlling for additional characteristics of prosecutors (age, age^2, $gender$, $Hierarchy$ - position in the system of prosecutors). Finally, Model 3 estimates a univariate OLS on a sample including a number of participants who did not identify the correct subsection.

$$Sentence = \alpha + \beta Treatment + \epsilon \quad \text{(Model 1)}$$

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

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

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

$$\text{Sentence} = \beta_1 \text{Scenario A} + \beta_2 \text{Scenario B} + \beta_3 \text{Scenario C} + \varepsilon \quad (1)$$

and using the Wald test, we test

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

We do not reject the null hypothesis suggesting that the 50k threshold has the same effect on the absolute length of incarceration as increasing the damage by CZK 435,000

Table 2: The Effect of Quantity Thresholds on Sentence Decisions for Theft

Panel I: 50k Threshold

	Model 1	Model 2	Model 3
Treatment Effect	10.153***	11.631***	10.056***
	(1.614)	(1.683)	(1.595)
Constant	18.826***	11.730**	18.826***
	(0.866)	(5.402)	(0.866)
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 the 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 a univariate OLS that includes responses of participants who did not identify the correct subsection of the paragraph.

(by 855%). Despite the probable sensitivity of the results on particular specifications and circumstances of cases, we take the results as robust evidence that the effect of thresholds on the shape of sentencing decisions is of great importance.

3.5 Drug Possession

3.5.1 Vignette description

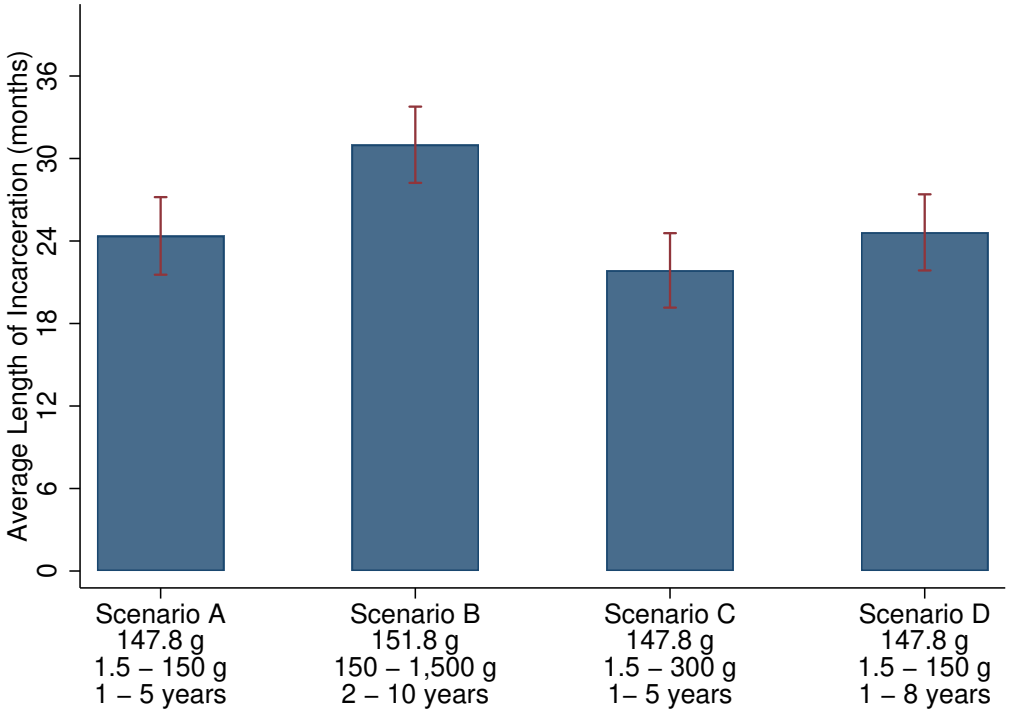
In the drug possession case, an offender was selling methamphetamine in front of a dance club in a town in Northern Bohemia. It was a one-time event after the offender lost his job. He had been repeatedly sentenced for selling marijuana over the previous 10 years and for small thefts. In four scenarios, we manipulate: (i) the amount of pure substance of methamphetamine (henceforth only methamphetamine) found on him; (ii) the applicable sentencing range; and (iii) the thresholds influencing the composition of cases within corresponding subsections. The full text of the vignette, along with the text of the section and other information provided to the participants, is presented both in English and in Czech in the Appendix. The four scenarios that were applied are summarized in Table 3. Scenario A and scenario B are based on the existing legislation and differ marginally in the amount of the drug possessed. The difference between the length of incarceration recommended in scenario A and B represents the overall threshold effect caused by the threshold.

Table 3: Scenarios of Drug Possession Offenses

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

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

Figure 4: Drug Possession: Average Recommended Sentence by Prosecutors



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

3.5.2 Results

Figure 4 shows the average sentence recommended by prosecutors in different scenarios. In scenario A, the average length of incarceration is slightly more than 2 years and in scenario B longer by more than 6 months. The figure also reveals that the lowest average sentence is in scenario C, slightly more than 22 months, and finally, scenario D is almost identical to scenario A.

Table 4 shows results from formal tests of three hypotheses; Panel I tests the overall threshold effect, Panel II tests for the reference mechanism, and Panel III tests for the severity mechanism. Similarly to the first case, each panel shows results from three empirical models: (i) a simple univariate OLS regression – conveniently, in all three panels the treatment dummy equals to zero for observations from scenario A; (ii) an OLS model controlling for additional characteristics of prosecutors – for the formal regression see equation Model 1; and (iii) a univariate OLS on a sample including several participants who did not identify the correct subsection.

Panel I of Table 4 provides convincing evidence that the threshold increases the length of incarceration by around 6 months, which represents approximately a 25% increase

of the length. Employing different models, the point estimates of the treatment effect range from 5.2 to 6.6 months. The results thus provide robust evidence that prosecutors recommended a significantly different length of incarceration for two almost identical cases (the difference is 3 grams of methamphetamine (2.7%)) .

Panel II of Table 4 shows suggestive evidence of the reference mechanism. All three specifications provide similar negative point estimates, suggesting that by increasing the threshold of the classifying variable from 150g to 300g, the sentence decreases by 2.5 - 3 months. The effect represents approximately a 10% decrease in the length of incarceration. Since only one point estimate is statistically significant, we consider the results only as suggestive evidence.

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

4 Measure of Justice

4.1 Just Sentence and Its Measure

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

We first define the *just* sentence formally. Using the notation introduced earlier,

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

Table 4: The Effect of Quantity Thresholds on Sentence Decisions for Drug Possession

Panel I: Effect of Threshold

	Model 1	Model 2	Model 3
Treatment Effect	6.629***	5.870***	5.240**
	(1.969)	(2.229)	(2.016)
Constant	24.370***	41.876	25.379***
	(1.407)	(25.440)	(1.449)
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 a univariate OLS that includes responses of participants who did not identify the correct subsection of the paragraph.

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

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

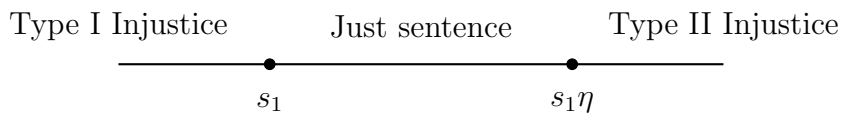
$$s_1 \leq s_2 \tag{D 2.1}$$

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

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

Definition 2 implies that a sentence that is not considered *just* can be of two types. Depending on which condition is not satisfied, we distinguish two types of *unjust* sentences. If condition D 2.1 is not satisfied and the more severe case leads to a more lenient sentence, then we refer to *Type I injustice*, whereas if condition D 2.2 is not satisfied and the sentence s_2 is too harsh, we refer to *Type II injustice*. Figure 5 graphically represents the idea.

Figure 5: Just Sentence



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

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

of CZK 48,283). There are also N^B decisions in cases $B = (x + \delta_x, t + \delta_t)$ denoted as s_j^B (e.g. a theft case with damage of CZK 51,283).

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

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

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

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

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

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

4.2 Measure of Justice in Experiment

We calculate the introduced measures for sentences recommended by prosecutors in our experiment. In particular, we focus on the threshold effect in the drug possession case, the 50k and the 500k thresholds effects in the theft case, and the size effect in the theft case. For each of these comparisons, we calculate *Type I injustice* and two measures of

justice with tolerance parameters of 1.5 and 2. Finally, we report an approximation of how tolerant the society must be to evaluate at least half of the sentences as *just*.

Table 5: Measures of (In)Justice in the Experiment

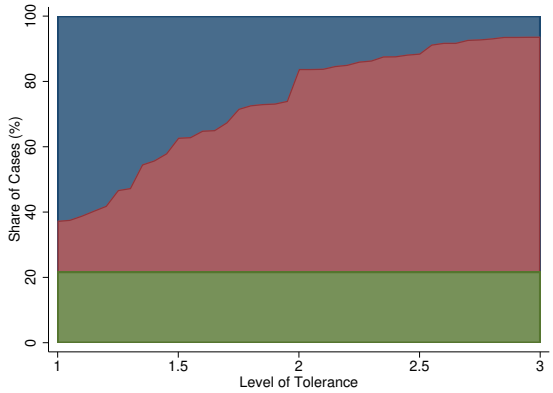
	Type I injustice	$\mathcal{M}^J(1.5)$	$\mathcal{M}^J(2)$	$\left(\mathcal{M}^J\right)^{-1}(0.5)$
Drug Possession	0.21	0.41	0.62	1.75
Theft (50k threshold)	0.15	0.38	0.62	1.85
Theft (500k threshold)	0.33	0.49	0.63	1.55
Size of Damage	0.15	0.41	0.65	1.70

Table 5 shows the results. The *Type I injustice* is highest in the 500k threshold comparison, as a third of the more severe cases were sentenced to a more lenient sentence. This suggests that these two cases (theft cases with damage of CZK 487,092 and CZK 508,213) are perceived the most similar among all four pairwise comparisons. The second column implies that if the society tolerates that a more severe case is sentenced to a 50% longer incarceration, then the share of *just* sentences is around 40% for most of the comparisons, with an exception for the 500k threshold. Once the tolerance parameter is two, roughly two thirds of cases are considered fair in all four comparisons. Note, however, that the interpretation differs. For example the 50k threshold effect and the 500k threshold effects differ dramatically in the measure of *Type I injustice*, and since the measures sum to 1 for any η, they also differ in the *Type II injustice*. Finally, an inverse function of *measure of justice* evaluated at 0.5 indicates how tolerant one has to be (i.e. what the η must be) to reach 50% probability that, from a randomly observed pair of sentences s_i^A and s_j^B, the latter is considered a just sentence. For example, the first row that compares a drug possession case with 148.8 grams and a drug possession case with 151.8 grams reveals that to reach 50% of just decisions, one has to consider a 75% longer incarceration for 3 grams of methamphetamine as a just sentence. This is additional evidence that the threshold has an enormous effect not only on the average sentence recommended. Figure 6 shows measure of justice and injustice for four main comparisons.

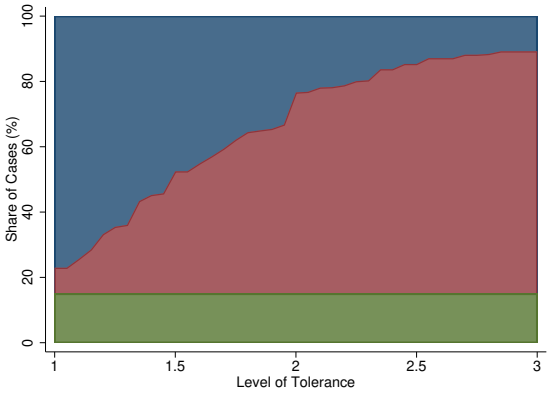
5 Discussion

Introducing a measure providing guidance to sentencers necessarily includes a trade-off. Measures envisioned to reduce sentencing disparities will likely introduce a certain amount of new ones. As we show, the extent of new disparities may be substantial and the disparities introduced by dividing offenses into subsections with specific sentencing

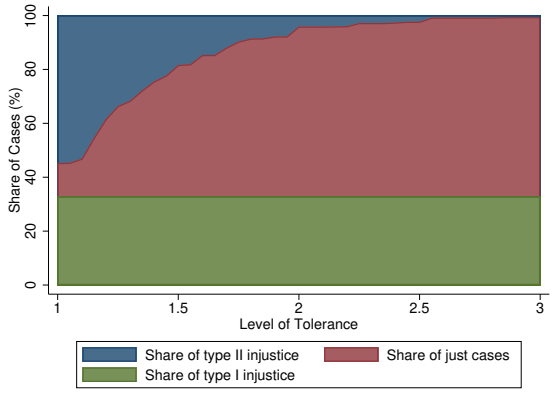
Figure 6: Measures of (In)Justice in Experiment



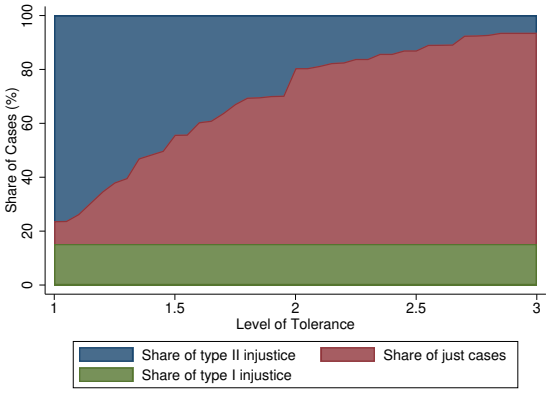
(a) Panel A: Drug Possession



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



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



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

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

ranges may strongly limit the benefits of such a measure. The difference in sentences imposed for identical thefts around a threshold (CZK 48,283 and CZK 51,283 scenarios) was as large as the difference in sentences imposed for identical thefts at the opposite sides of one offense subsection (CZK 51,283 and CZK 487,092). In other words, stealing more by CZK 3,000 leads to a similar increase in sentence as stealing more by CZK 435,000. Any criminal justice system striving for principled sentencing cannot tolerate such a disparity.

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

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

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

limits. Since most of the sentences seem to be clustered in the lower part of the sentencing ranges close to the lower limit (Drápal, 2020; Sutela, 2020), increasing the lower limit would lead to a higher increase in the average sentence.

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

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

Studying the average effect masks substantial heterogeneity in prosecutors' behavior. Figure 7 and Figure 8 in the Appendix show histograms of all sentences recommended by individual prosecutors for each scenario. The distribution of recommended sentences within each scenario exhibits a large variation. Strikingly, in the majority of scenarios, they chose sentences close to both the lower and the upper limits of sentencing ranges. Furthermore, the figures show that prosecutors tend to round the length to years, which is in line with previous research by Dhami et al. (2020); Pease and Sampson (1977). Quantity thresholds are thus not the only source of sizable sentencing disparities in the Czech criminal justice system.

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

We further propose a novel measure of justice. We formalize the notion of justice such that similar cases should be treated similarly and different cases differently and introduce a formal definition of a *just* sentence. In particular, a sentence is considered

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

The structure of offense subsections with sentencing ranges substantially shapes sentencing. We close by remarking that there is a lack of both theoretical and empirical scholarship regarding how the subsections and ranges should be constructed. What classifying variables should divide offense subsections, into how many subsections, and should some classifying variables be quantifiable? To what extent should sentencing ranges for individual subsections overlap? Providing answers to these questions seems necessary to understand how offense subsections should influence sentencers.

References

J. M. Anderson, J. R. Kling, and K. Stith. Measuring Interjudge Sentencing Disparity: Before and After the Federal Sentencing Guidelines. *The Journal of Law and Economics*, 42(S1):271–308, apr 1999. ISSN 0022-2186. doi: 10.1086/467426. URL `https://www.journals.uchicago.edu/doi/10.1086/467426`.

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

D. Bjerk. Making the Crime Fit the Penalty: The Role of Prosecutorial Discretion under Mandatory Minimum Sentencing. *The Journal of Law and Economics*, 48(2):591–625, oct 2005. ISSN 0022-2186. doi: 10.1086/429980.

D. Bjerk. Mandatory Minimums and the Sentencing of Federal Drug Crimes. *The Journal of Legal Studies*, 46(1):93–128, jan 2017a. ISSN 0047-2530. doi: 10.1086/690205.

D. Bjerk. Mandatory Minimum Policy Reform and the Sentencing of Crack Cocaine Defendants: An Analysis of the Fair Sentencing Act. *Journal of Empirical Legal Studies*, 14(2):370–396, jun 2017b. ISSN 17401453. doi: 10.1111/jels.12150. URL `http://doi.wiley.com/10.1111/jels.12150`.

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

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

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

M. K. Dhami, I. K. Belton, E. Merrall, A. McGrath, and S. M. Bird. Criminal Sentencing by Preferred Numbers. *Journal of Empirical Legal Studies*, 17(1):139–163, mar 2020. ISSN 1740-1453. doi: 10.1111/jels.12246. URL `https://onlinelibrary.wiley.com/doi/abs/10.1111/jels.12246`.

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

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

J. Fleetwood. Five Kilos: Penalties and Practice in the International Cocaine Trade. *British Journal of Criminology*, 51(2):375–393, mar 2011. ISSN 0007-0955. doi: 10.1093/bjc/azr006.

J. A. Foulds and D. Nutt. Principled Sentencing for Drug Supply Offences: Revised Methamphetamine Sentencing guidelines in New Zealand. *Drug Science, Policy and Law*, 6:205032452094234, jan 2020. ISSN 2050-3245. doi: 10.1177/2050324520942347. URL http://journals.sagepub.com/doi/10.1177/2050324520942347.

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

P. J. Hofer. Federal Sentencing after Booker. *Crime and Justice*, 48:137–186, may 2019. ISSN 0192-3234. doi: 10.1086/701712. URL https://www.journals.uchicago.edu/doi/10.1086/701712.

B. D. Johnson. Trials and Tribulations: The Trial Tax and the Process of Punishment. *Crime and Justice*, 48:313–363, may 2019. ISSN 0192-3234. doi: 10.1086/701713. URL https://www.journals.uchicago.edu/doi/10.1086/701713.

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

R. Kert, B. Slot, G. Vernimmen, A. Lehner, and J. Schellekens. Study on minimum sanctions in the EU Member States. Technical report, European Commission, 2015. URL https://op.europa.eu/en/publication-detail/-/publication/1226bed2-be78-11e5-9e54-01aa75ed71a1/language-en.

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

L.-P. Lepage. Are Criminals Strategic? Offender Responses to Drug Sentencing Cutoffs. *Labour Economics*, 66:101906, oct 2020. ISSN 09275371. doi: 10.1016/j.labeco.2020.101906. URL https://linkinghub.elsevier.com/retrieve/pii/S092753712030110X.

K. Pease and M. Sampson. Doing Time and Marking Time. *The Howard Journal of Criminal Justice*, 16(2):59–64, oct 1977. ISSN 02655527. doi: 10.1111/j.1468-2311.1977.tb00333.x. URL http://doi.wiley.com/10.1111/j.1468-2311.1977.tb00333.x.

Sentencing Council. Drug Offences Guideline: Professional Consultation. Technical report, 2011. URL https://www.sentencingcouncil.org.uk/wp-content/uploads/Drug_Offences_Guideline_Professional_Consultation.pdf.

L. O. Shermer and B. D. Johnson. Criminal Prosecutions: Examining Prosecutorial Discretion and Charge Reductions in U.S. Federal District Courts. *Justice Quarterly*, 27(3):394–430, jun 2010. ISSN 0741-8825. doi: 10.1080/07418820902856972. URL https://www.tandfonline.com/doi/full/10.1080/07418820902856972.

D. Skugarevskiy. *Essays in law and economics of enforcement.* Dissertation thesis, Graduate institute of international and development studies in Geneva, 2017. URL https://repository.graduateinstitute.ch/record/295661/files/thesis_skougarevskiy_30nov2017_electronic_submission.pdf.

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

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

M. Tonry. Fifty Years of American Sentencing Reform: Nine Lessons. *Crime and Justice*, 48:1–34, may 2019. ISSN 0192-3234. doi: 10.1086/701798. URL https://www.journals.uchicago.edu/doi/10.1086/701798.

E. Travova. Under pressure? performance evaluation of police officers as an incentive to cheat: Evidence from drug crimes in russia. *CERGE-EI Working Paper Series*, (637), 2019.

C. Tuttle. Racial Disparities in Federal Sentencing: Evidence from Drug Mandatory Minimums. 2019. URL http://econweb.umd.edu/$\sim$tuttle/files/tuttle_mandatory_minimums.pdf.

J. T. Ulmer. Recent Developments and New Directions in Sentencing Research. *Justice Quarterly*, 29(1):1–40, feb 2012. ISSN 0741-8825. doi: 10.1080/07418825.2011.624115.

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

United States Sentencing Commission. 2011 Report to the Congress: Mandatory Minimum Penalties in the Federal Criminal Justice System. Technical report, 2011. URL https://www.ussc.gov/research/congressional-reports/2011-report-congress-mandatory-minimum-penalties-federal-criminal-justice-system.

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

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

Appendix

Proof of Theorem 1

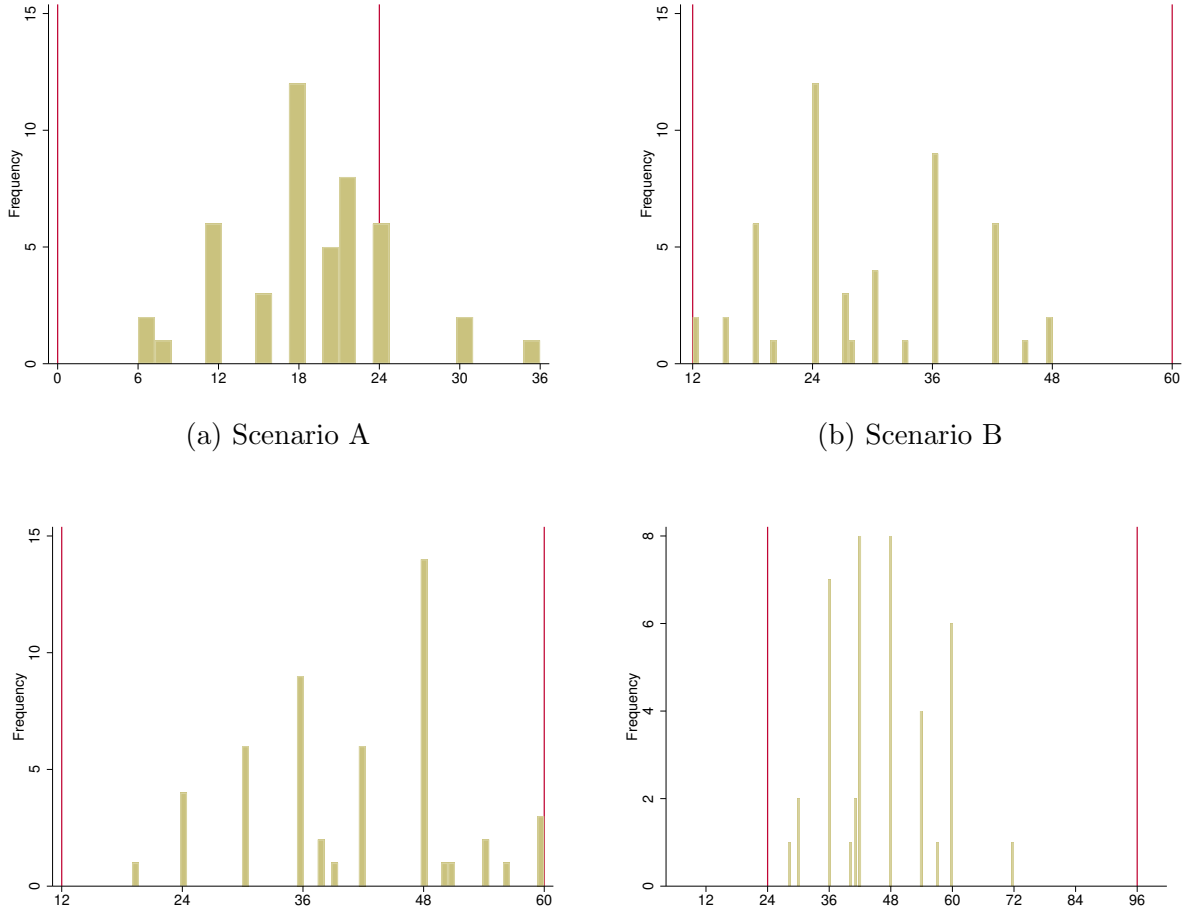
We show that equation T 1.1 holds. To simplify notation, label two cases A and B and define three differences $\Delta G = G(B) - G(A)$, $\Delta\rho^- = \rho^-(B) - \rho^-(A)$, and $\Delta\rho^+ = \rho^+(B) - \rho^+(A)$. Then, according to Definition 1, the difference in sentences $\Delta s = s(B) - s(A)$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

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

□

6 Appendix B

Figure 7: Theft: Individual Sentences Recommended by Prosecutors



(a) Scenario A

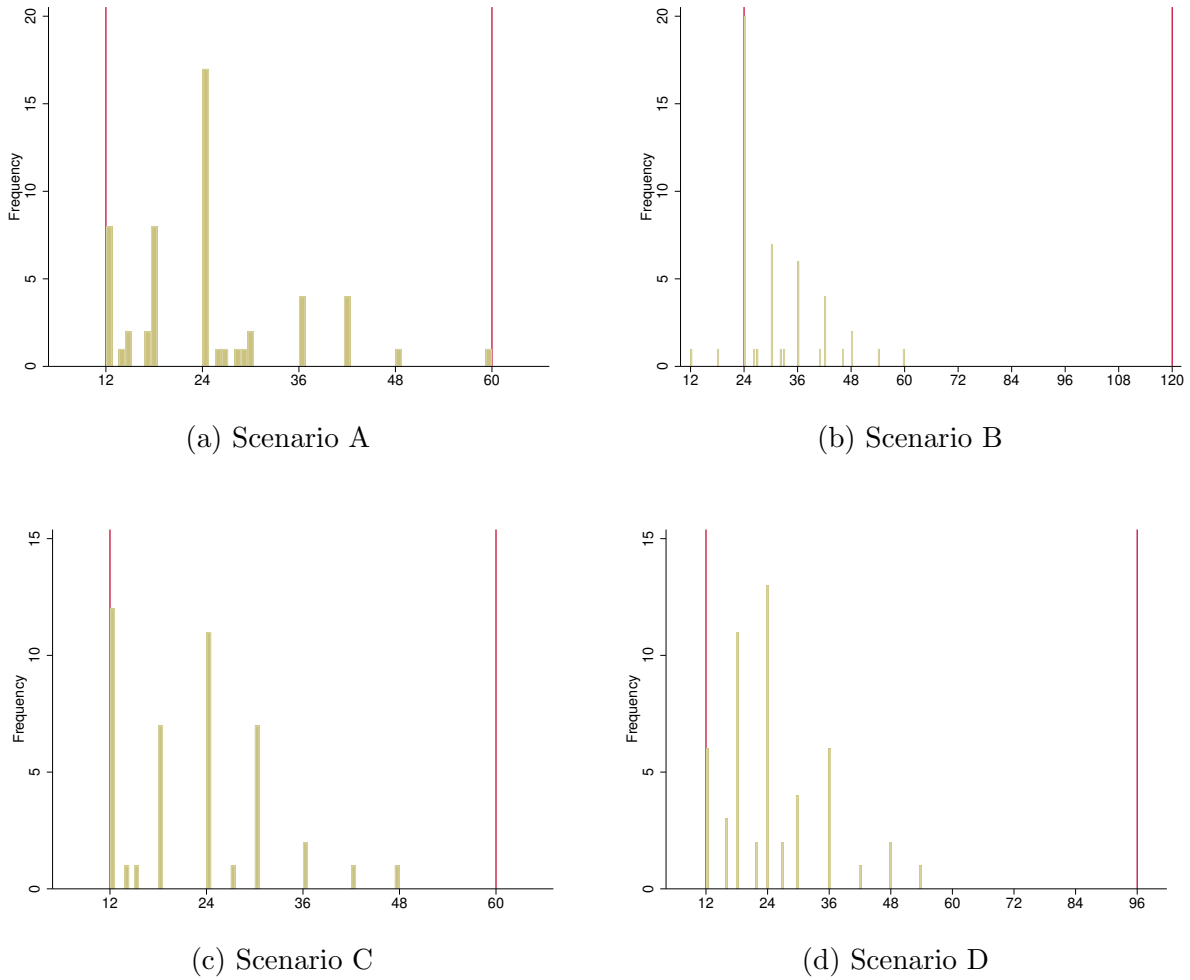
(b) Scenario B

(c) Scenario C

(d) Scenario D

Notes: Each panel represents a histogram of individual recommended sentences in the case of theft. In scenario A, prosecutors recommended sentences in a case with damage of CZK 48,283. In scenario B, the prosecutors recommended sentences in a case with damage of CZK 51,283. In scenario C, prosecutors recommended sentences in a case with damage of CZK 487,092. Finally, in scenario D, prosecutors recommended sentences in a case with damage of CZK 508,213. Red vertical lines mark the upper and lower limits of the corresponding sentencing range. See Table 1.

Figure 8: Drug Possession: Individual Sentences Recommended by Prosecutors



(a) Scenario A

(b) Scenario B

(c) Scenario C

(d) Scenario D

Notes: Each panel represents a histogram of individual recommended sentences in the drug possession case. In scenario A, prosecutors recommended sentences in a case with 147.8 g. In scenario B, the prosecutors recommended sentences in a case with 151.8g. In scenario C, prosecutors recommended sentences in a case with 147.8 g and a composition of more serious cases. Finally, in scenario D, prosecutors recommended sentences in a case with 147.8 g and higher upper limit of the sentencing range. Red vertical lines mark the upper and lower limits of the corresponding sentencing range. See Table 3.

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
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 the p-value of the two-sided t-test under the null that the sample value equals to the population (universe of all active prosecutors) value.

Introduction Screen[16] Welcome!
We ask you to participate in a scientific study, in which we explore sentencing decisions.

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

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

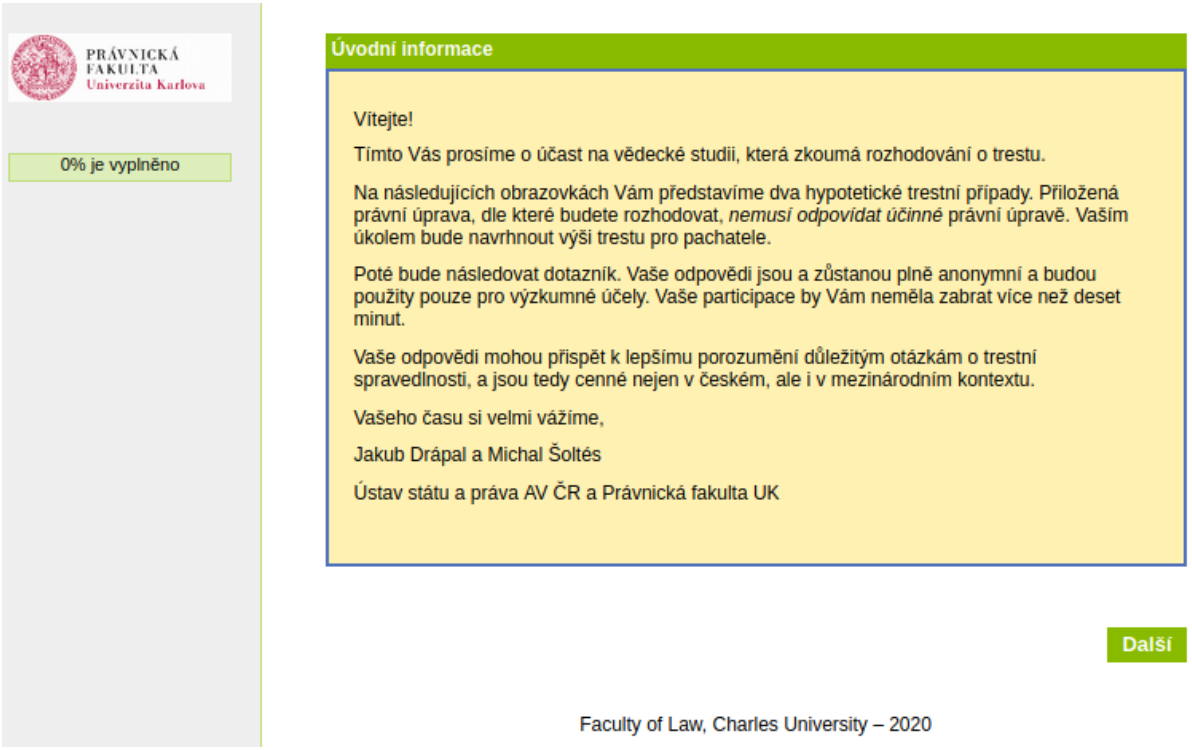
Your answers might contribute to the better understanding of important criminal justice issues that are significant not only for the Czech Republic, but for the international audience as well.

We highly value your time.

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

[16] See figure 9

Figure 9: Vignettes: Introduction Screen



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

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

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

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

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

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

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

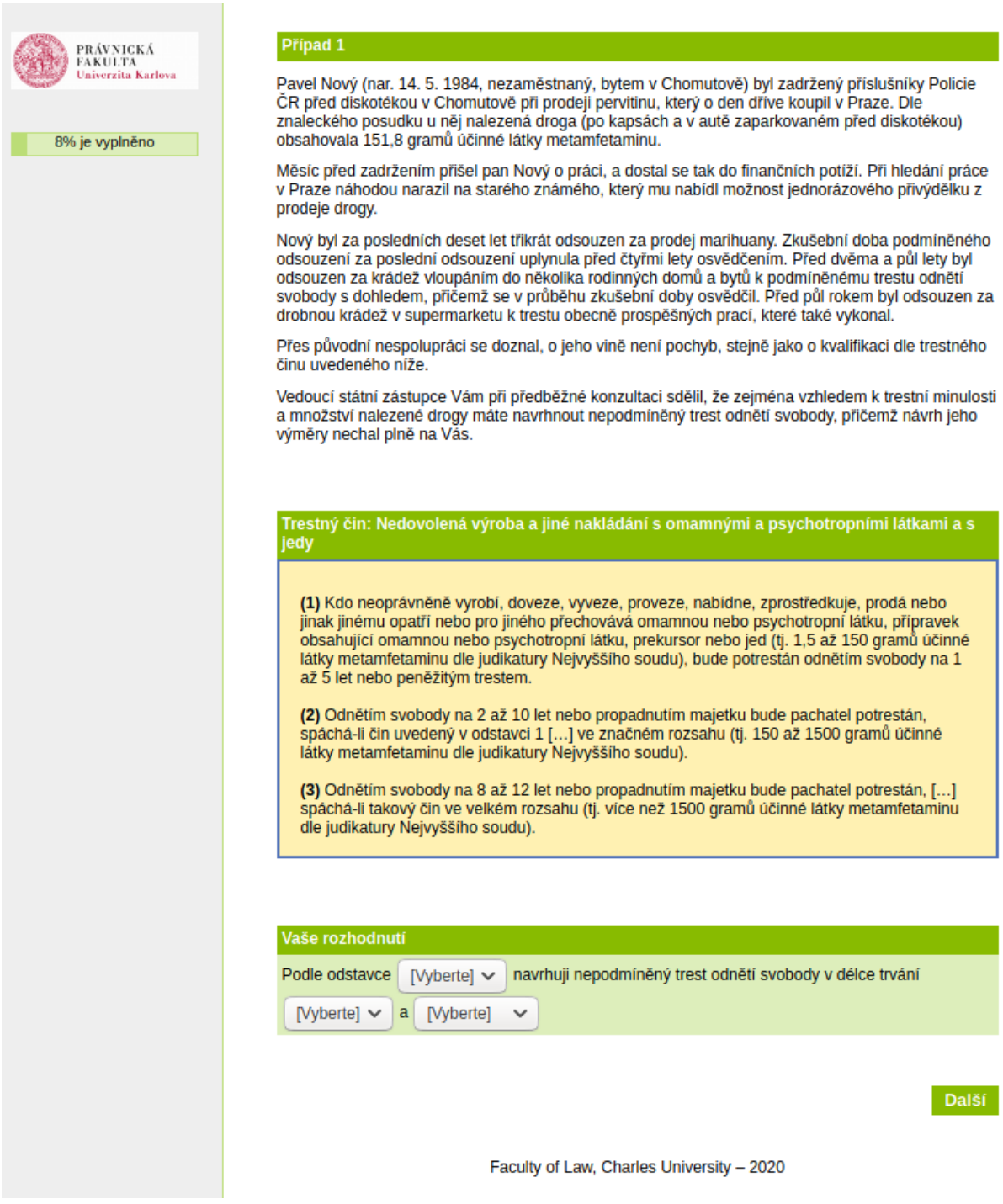
[17] See figure 10

(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 10: Vignette 1: Drug Distribution



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

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

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

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

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

Offense Section: Theft

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

...

(3) An offender shall be sentenced to imprisonment for one year to five years

[18] See figure11

or to a pecuniary penalty, if he/she causes larger damage (meaning CZK 50,000-500,000) by the act referred to in Sub-section (1).

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

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

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

Figure 11: Vignette 2: Theft