Working Paper Series 715
(ISSN 1211-3298)

Sentencing Decisions Around Quantity Thresholds: Theory and Experiment

Jakub Drápal
Michal Šoltés

CERGE-EI
Prague, December 2021

ISBN 978-80-7343-522-6 (Univerzita Karlova, Centrum pro ekonomický výzkum a doktorské studium)
ISBN 978-80-7344-617-8 (Národohospodářský ústav AV ČR, v. v. i.)

Sentencing Decisions Around Quantity Thresholds: Theory and Experiment*

Jakub Drápal† and Michal Šoltés‡

November, 2021

Abstract

We study the implications of the structure of criminal codes on sentencing decisions. To limit sentencing disparities, criminal codes typically divide offenses 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 drive substantial increases in sentences, 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 Michal Bauer, Libor Dušek, Francesco Drago, Christoph Engel, Brian Johnson, Ondřej Kubů, Filip Matějka, Josef Montag, Vladimír Novák, Jiří Pavlík, Ondřej Šťastný, Jan Zápal, conference participants at the GLEA 2021, EALE 2021, ERoS 2020, seminar participants at Erasmus University, Rotterdam and the Prosecutor General's Office and to prosecutors for their participation. This study was supported by Czech Grant Agency, project GA19-15077S "Sentencing disparities in the post-communist continental legal systems", 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

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

‡ CERGE-EI, a joint workplace of Center for Economic Research and Graduate Education, Charles University and the Economics Institute of the Czech Academy of Sciences, Politických vězňů 7, P.O. Box 882, 111 21 Prague 1, Czech Republic

1 Introduction

Scholars have been studying sentencing shortcomings since as early as 1972, when 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, backfired and became new sources of disparities. For example, while the US federal sentencing guidelines reduced the level of sentencing disparities at the court level (Anderson et al., 1999), they contributed to higher racial disparities (Hofer, 2019). 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 US federal sentencing guidelines need to be repudiated (Tonry, 2019). The US federal sentencing guidelines demonstrate how efforts to reduce sentencing disparities may unintentionally cause new disparities.

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

We argue that quantity thresholds are likely to introduce a new form of sentencing

[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.

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 sizeable sentencing disparities. Furthermore, the theory formalizes how the structure of criminal codes (e.g., overlaps of sentencing ranges) affects the probability and size of sentencing disparities caused by the quantity thresholds.

Figure 1: Theft Subsections with Sentencing Ranges

Notes: This figure represents the problem using an example of theft and the Czech criminal code. The offense of theft is divided into four subsections by quantity thresholds based on damage caused (CZK 50,000, CZK 500,000, and CZK 5,000,000; note that from October 2020, several months after our experiment, the thresholds doubled). For each subsection, there is a specific sentencing range that determines the upper and lower limits of the sentencing range i.e., length of incarceration. Note that, in many cases, two neighbouring sentencing ranges overlap and thus effectively permit so-called perverse sentencing when a (marginally) more serious case is sentenced with 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* mechanism and argue that it leads to harsher sentences. On the other hand, case B is compared to arguably more serious cases in its subsection. As a result, the relative position within the sentencing range is likely to be lower. We call this the *reference* mechanism and argue that it tends to decrease the sentence. Depending on which mechanism dominates, the effect of the threshold is either positive – case B is sentenced more harshly –, or negative – case B is sentenced less harshly. If the latter is the case, we refer to perverse sentencing, in which (marginally) more serious cases are sentenced to a more lenient punishment.

To provide empirical evidence, we conduct a Rachlinski-style online experiment with 200 Czech prosecutors. In the experiment, we focus on two crime cases in which quantity thresholds play a prominent role: theft and drug possession. In both cases, we implement several different scenarios in which we vary the amount of classifying variables (amount of methamphetamine possessed and damage caused) around thresholds. Each participant was randomly assigned to one scenario in the drug possession case and one in the theft case. Everyone was thus asked to recommend a sentence in two different crime cases. We then use the exogenous variation in amount of classifying variable to estimate the causal effects of the quantity thresholds on the recommended sentences. Since the participants are professional sentencers whose decisions and attitudes are crucial for final sentencing decisions, our results have immediate implications for current practice.

The causal effect of the quantity thresholds on sentencing decisions is substantial and leads to a sizeable increase in sentences. In the theft vignette, we vary the amount of damage in two scenarios around the thresholds of CZK 50,000 and two scenarios around the threshold of CZK 500,000 (henceforth the 50k and the 500k thresholds, respectively). To estimate the causal effect of the 50k threshold, we compare recommended sentences in two identical cases with marginally different damages 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 sentences in identical cases with damage of CZK 487,092 and CZK 508,213 and estimate the effect to be around 4 months, which corresponds to a 10% increase. 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. Specifically, we study the effect of a threshold defined as 150g of methamphetamine. 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.

To test for the existence of the severity and the reference mechanisms, we implement two additional scenarios in the drug possession case. These scenarios introduce an isolated variation in the composition of cases in an offense subsection and sentencing ranges, respectively. Due to institutional constrains, we could not implement scenarios that would generate the variation necessary to identify the full severity and reference effects. Instead, we only test for the mechanisms. First, we found suggestive evidence that the reference mechanism is negative as predicted. Second, we found no effect on the severity mechanism. However, since the overall effect is positive, our theory implies that the the severity effect must be positive as well.

Finally, to provide more comprehensive evidence on sentencing disparities and injustice caused by the quantity thresholds, we propose a novel parametric measure of sentencing (in)justice. Consider two cases A and B and corresponding sentences s^A and s^B and suppose that case B is marginally more serious. We define sentence s^B to be *just*, if it is not more lenient than sentence s^A and not too (unreasonably) harsh compared to sentence s^A. Consequently, there are two reasons why sentence s^B might not be considered just: (i) sentence s^B is more lenient than sentence s^A (*type I injustice*); and (ii) sentence s^B is too harsh (*type II injustice*). Using these definitions, we quantify the shares of just decisions in our experiment. The results generally follow those of the average recommended sentences. Due to the quantity thresholds, prosecutors view and treat almost identical cases differently. For example, to consider at least 50% of all sentencing decisions in the drug possession case as a *just sentence*, one has to tolerate that 3 additional grams of methamphetamine can increase the length of incarceration by 75%.

The rest of the paper is organized as follows. We first propose a theory that introduces the role of quantity thresholds on sentences, explains mechanisms, and guides our 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 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 CZK 49k (case A) and CZK 51k (case B) – fall in two subsections with different sentencing ranges, resulting in possibly different sentences for case A and case B. Sentencing ranges serve as a rudimentary signpost indicating to what extent severe sanctions are expected. The composition of cases within a corresponding subsection provides a natural reference group, within which cases are

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

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

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

[6]Sections 153 and 159.

compared with each other and are ordered by seriousness (this extends the statistical curving proposed by Leibovitch (2016a)). 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 sentences in case B and case A caused by different sentencing ranges as the *severity* effect. On the other hand, case B is also compared to arguably more serious cases within its subsection and thus the relative position within the sentencing range is lower. We call the difference between sentences in case B and case 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, as it provides new guidance regarding the cardinal ranking of offenses (Leibovitch, 2016a). Thus, if a sentencing range for a specific offense subsection is increased (either the lower or the upper limit is increased), the length of sentences imposed are likely to increase as well.

Conversely, if only the composition of cases within offense subsections changes – more or less severe cases are added or removed – the relative position of a specific case is transposed within the offense subsection. Alternatively, the same case transposition within the subsection can be evoked by a contrast effect that changes sentencers' perception of the seriousness of the original case (Leibovitch, 2016b). 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 for two marginally different cases around quantity thresholds generally affect sentencing consideration in opposite directions. Depending on which effect dominates, the quantity threshold either increases the sentence (in case B a harsher sentence is imposed) or decreases it (in case B a more lenient sentence is imposed). 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 to several studies on the United States and on Russia. Using observable data from the United States, Bjerk (2017b) documents that judges imposed different sentences to offenders with a drug amount 10% above and below quantity thresholds. However, these effects became smaller or disappeared completely once he controlled for other observed characteristics such as weapons charge. Studying the threshold for 10 year mandatory minimums, Tuttle (2019) finds an important increase upon crossing the threshold. The effect was largely driven by prosecutorial discretion. Thresholds of composite scores for seriousness and criminal history have also been found to influence sentencers (Pintoff, 2004).

The effects of mandatory minimums cannot, however, be easily generalized to lower limits of sentencing ranges. High mandatory minimums (5 and 10 years) and large prosecutorial discretion lead to a bunching of cases just below quantity thresholds (Bjerk, 2017b, 2005; Tuttle, 2019). Furthermore, most criminal justice systems around the world have less developed plea-bargaining systems than the United States (Johnson, 2019); the prosecution has less power and, as a result, it cannot similarly influence sentencing outcomes around thresholds. Additionally, mandatory minimums are often not applied even during sentencing; note that they are imposed e.g. upon fewer than half of drug offenders who were eligible for mandatory minimums at sentencing (United States Sentencing Commission, 2011, Chapter 8). Altogether, in systems other than in the US, prosecutors, offenders, and judges have fewer opportunities and less power to place the case above or below the threshold, to impose sentences below the lower limit of sentencing ranges, and they have less incentive to do so.

In a study most relevant to ours, Skugarevskiy (2017) uses administrative data and examines sentencing for drug offenses in Russia. He finds that crossing a threshold of 100 grams for cannabis and 2.5 grams for heroin leads to an increase of 0.84 years of imprisonment. Note that the Russian setting is a specific case of the setting we study, as the 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. Therefore, our study is the first to investigate both the role of (continental-style) thresholds in sentencing and of the underlying mechanisms at play.

2.2 Conceptual Framework

We next introduce a conceptual framework that formalizes the previous discussion on the role of thresholds in sentencing decisions. We assume that for a given offense, criminal codes specify a classifying variable, quantity thresholds $\tau \in \mathcal{T} = \{\tau^{[0]}, \tau^{[1]}, \tau^{[2]}, \ldots\}$ that split the offense into subsections, and corresponding sentencing ranges i.e., intervals $\rho(\tau) = (\rho^{-}(\tau), \rho^{+}(\tau))$ that restrict the space for a possible sentence.[7] We further assume that any case can be fully characterized by a pair of (x, t), where x represents all relevant factors of the case and t is the amount of the classifying variable (e.g., amount of drugs possessed).

A sentencing process is a two-stage rule whereby any offense (x, t) is assigned a sentence s. In the first stage, the rule classifies an offense (x, t) into a corresponding subsection $\tilde{\tau}$ by comparing the amount 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. In particular, we assume that the final sentence s is a linear combination of the lower and the upper limits of the corresponding sentencing range with relative seriousness $G(x, t; q(\tilde{\tau}))$ as a weight.

Definition 1 (Sentencing Rule)**.** A sentence s imposed for an offense (x, t) is determined by the following two-step *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 a case (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*. We assume that the recommended sentence must be within the corresponding sentencing range i.e., 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 a case increase, so does the case'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

[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.

will be positioned lower in the sentencing range. Assumption 1 introduces the properties formally.

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

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

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

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

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,

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

[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.

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 opposing 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.

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 in both subsections 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.

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.

[9] See Appendix A for more details.
[10] In fact, the *severity effect* is likely to be positive.

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.

2.2.1 Implications

The framework allows us to discuss how the structure of the subsections (i.e., overlap of sentencing ranges) shapes sentencing more generally. In some countries, such as in France, the sentencing ranges are organized so the lower limits of the sentencing range are 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 that the thresholds are identical in both systems ($\mathcal{T}^\alpha = \mathcal{T}^\beta$). For illustrative purposes, consider a situation with only two quantity thresholds τ^0 and τ^1. 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 upper limit of the sentencing range increases. Additionally, we assume that for τ^0 both systems have the same lower limits of the sentencing range. Therefore, $\rho_\alpha^-(\tau^0) = \rho_\alpha^-(\tau^1) = \rho_\beta^-(\tau^0) < \rho_\beta^-(\tau^1)$. Figure 2 represents the structure of both systems. Note that neither of the systems violates our Assumption 2.

Our framework implies that the severity effect is weaker in the α (French-like) system. This follows from the fact that $\Delta\rho_\alpha^- = 0$ and the first term of the severity effect stated in Corollary 1 is thus nullified. Consequently, the expected difference between sentencing decisions in two marginally different cases around the threshold is lower. Additionally, the probability of a perverse sentence is higher in the α than in the β system.[11]

Whether and how fast the lower limits of sentencing ranges grow with a higher sub-

[11]This comparative statics assumes that the $G(x, t; q(\tau))$ is the same in both systems.

section 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. We benefit from conducting the experiment with prosecutors who who are largely responsible for the final sentences in criminal cases. Using experienced and professional sentencers limits the concerns that the results are driven by a lack of awareness and experience with the criminal code and sentencing decisions. Furthermore, our results have immediate implications for current practice in sentencing decisions. We provide empirical evidence from two offenses with a prominent role of quantity thresholds: theft and drug possession.

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 sentences imposed, documented by a high increase in the fine imposition rate resulting from intentional effort of prosecutors in 2016 (Drápal and Dušek, 2021).

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 for indeterminate terms 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.

3.2 Implementation of the Experiment

3.2.1 Experimental Subjects

To approach prosecutors, we partnered with the Prosecutor General's Office of the Czech Republic. The invitation to participate in the online 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]

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 Appendix B.

3.2.2 Experimental Design

To estimate the causal effects of thresholds on the average recommended sentence, we manipulate the amounts of the classifying variables around the thresholds. We randomized prosecutors into several different scenarios (treatment arms). In each scenario, participants were instructed to analyze a criminal case and recommend a length of incarceration. The cases presented were identical except for the amounts of the classifying variables, which we manipulated to create marginally different cases around the quantity thresholds. This design allows us to explore the between-subjects variation and use the experimental manipulation to estimate a causal effect of the threshold effect. Note that the estimated differences in average recommenced sentences capture both the threshold effect and a slight increase in the classifying variable. Since the effect of an increase in a classifying variable on the final sentence is negligible, we attribute the all change to the threshold effects.

Furthermore, we test for severity and reference effects. Theorem 1 implies that testing for the severity and reference effects requires the other mechanism to be muted. To test for the severity effect, the reference effect must be suppressed and *vice versa*. Since this is not possible under the existing criminal code, we introduce two scenarios with hypothetical criminal codes varying sentencing ranges and offense compositions, respectively. To test for the reference effect, we increased the quantity threshold and thus effectively added more severe cases into the subsection. Similarly, to test for the severity effect, we implement scenario with increased upper limits of the sentencing range. To minimize inconsistency in the criminal code in these hypothetical situations, we rely only on small changes. 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.

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

The scenarios with hypothetical criminal codes were implemented in the drug possession case.

Each participant was randomized into one scenario with theft case and one scenario with drug possession case.[13] The second randomization was independent from the first one.[14] Each criminal 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.

3.3 Theft

3.3.1 Vignette description

Prosecutors were asked to recommend a length of incarceration in the following criminal 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

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

[14] Note that, because not everyone who started the survey also finished it, the randomization based on entry does not necessarily lead to a numerically identical number of observations in each scenario.

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 Appendix C. Table 1 shows specifications of the four scenarios.[15] We are interested in three causal effects. 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, note that scenarios B and C differ only in the size of the damage (i.e., classifying variable), but both cases are sentenced according to the same subsection with the same sentencing range. Therefore, the comparison between the average recommended sentence estimates the causal effect of the increase in damage caused.

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

Notes: This table summarizes four scenarios in the theft case. In all four scenarios, the prosecutors are instructed to apply the existing criminal code. The cases are identical except for the damage caused. In scenario A and B, we manipulate the damage so it is just below and just above the 50k threshold. In scenarios C and D, we manipulate the damage so it is just below and just above the 500k threshold.

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.

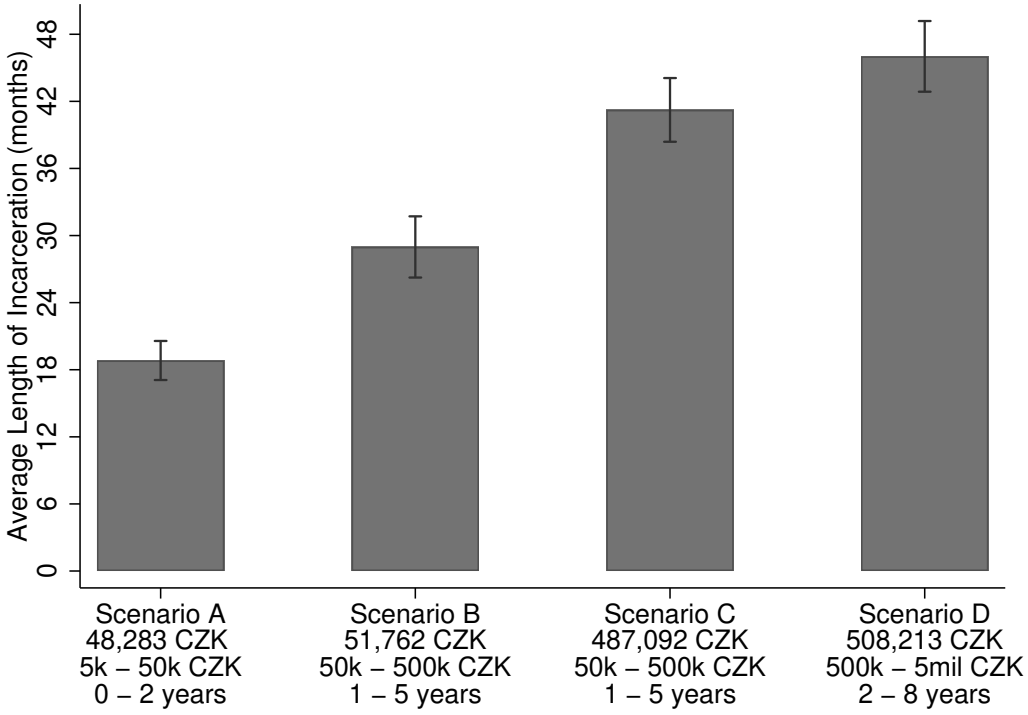
3.3.2 Results

Figure 3 shows the average length of incarceration recommended by prosecutors in different scenarios.[16] The higher the damage caused, the longer the recommended sentence.

[15] 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.

[16] Figure 7 in Appendix B shows the distribution of recommended sentences for each scenario.

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.

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 report results from three models. Model 1 represents a simple univariate OLS regression in which the treatment effect is captured by a dummy that equals to 1 if the observation comes from the treated scenario and 0 otherwise. See equation in Model Model 1. Model 2 extends the univariate OLS by controlling for additional characteristics of prosecutors X (Age, $Male$, $Alma Mater$, $Communist Party$, $Tenure Oath$, $Pros. Office$ - position in the system of prosecutors). Finally, Model 3 estimates a univariate OLS on a sample of participants from the district prosecutor's office. Since district prosecutors deal with cases of equivalent seriousness on a daily basis, their results are highly policy relevant.

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

$$Sentence = \alpha + \beta Treatment + \delta X + \epsilon \quad \text{(Model 2)}$$

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 damage. Panel I presents robust evidence that the 50k threshold increases the average incarceration by about 10 months, which represents more than a 54% increase compared to the sentence for cases just below the 50k threshold. When we control for other characteristics of prosecutors, the point estimate is even larger. The effect among district prosecutors is 7.6 months, but still statistically significant. 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%). Model 3 suggests that the effect among district prosecutors is even slightly higher (5.4 months). While the effect of the 500k threshold is generally smaller than the one caused by the 50k threshold, it provides additional evidence that the existence of the threshold effect is rather universal 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 in all three specifications. Table 7 in Appendix B reports full results from Model B.

We next re-estimate Model B allowing for a gender specific and an office specific treatment effect, respectively. Interestingly, the 500k threshold effect seems to be driven by female prosecutors, who increase the recommended length of incarceration by 10 months. There is no statistically significant effect among male prosecutors. Due to the limited number of observations, the heterogeneous treatment effects must be interpreted with caution. See Table 7 for the full results.

The effect of the size of damage allows us to understand the enormous effect the 50k threshold. Increasing a damage by CZK 435,000 (by 855%) corresponds to additional 12 months of incarceration. If we were to impose a linear projection, one month of incarceration corresponds to damages of CZK 36,000.[17] This is in contrast to the 10 month increase in incarceration for an increase in damage by CZK 3,000 estimated in Panel I. We consider this to be supportive evidence of why the differences in average recommended sentences can be attributed to the threshold effect rather than slight increases in the damage caused.

Alternatively, we can compare the 50k threshold effect to the effect caused by increas-

[17]We use the linear relationship only for demonstrational purposes. We do not argue that the sanction should increase in damage linearly.

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***	11.570***	7.677***
	(1.614)	(1.736)	(1.816)
Constant	18.826***	18.159*	19.906***
	(0.866)	(9.467)	(1.013)
Control Variables	No	Yes	No
Prosecutors	All	All	District Off.
N	96	91	68

Panel II: 500k Threshold

	Model 1	Model 2	Model 3
Treatment Effect	4.789**	4.676*	5.472**
	(2.110)	(2.430)	(2.433)
Constant	41.235***	34.879***	39.297***
	(1.420)	(10.054)	(1.461)
Control Variables	No	Yes	No
Prosecutors	All	All	District Off.
N	92	85	63

Panel III: Size of Damage

	Model 1	Model 2	Model 3
Treatment Effect	12.255***	11.651***	11.713***
	(1.966)	(2.263)	(2.096)
Constant	28.98***	31.912***	27.583***
	(1.361)	(11.176)	(1.504)
Control Variables	No	Yes	No
Prosecutors	All	All	District Off.
N	101	96	73

Robust standard errors in parentheses.

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

Notes: Panel I tests the 50k threshold effect (s(B) - s(A)), Panel II tests the 500k threshold effect (s(D) - s(C)), and Panel III tests the effect of the size of damage (s(C) - s(B)). *Treatment* corresponds to a dummy variable which equals 1 for respondents who were assigned to the treatment group with a higher damage. Model 1 represents a simple univariate OLS regression, Model 2 extends the univariate OLS by controlling for additional characteristics of prosecutors. For full results see Table 7 in Appendix B. Model 3 is a univariate OLS on a sample of district prosecutors. Prosecutors who failed our attention check are dropped in all three specifications.

ing the damage by CZK 435,000. Formally, we run the following regression

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

where Scenario A, Scenario B, and Scenario C are indicators which equal 1 if the observation comes from the corresponding scenario. We then apply the Wald test to test

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

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

3.4 Drug Possession

3.4.1 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 Appendix C. 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.

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

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

Notes: This table summarizes four scenarios in the drug possession case. In scenario A and B, we manipulate the amount of methamphetamine possessed so it is just below and just above the 150g threshold. In scenarios C and D, we introduce hypothetical criminal codes to manipulate the structure of cases in the corresponding subsections and the sentencing ranges.

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.

3.4.2 Results

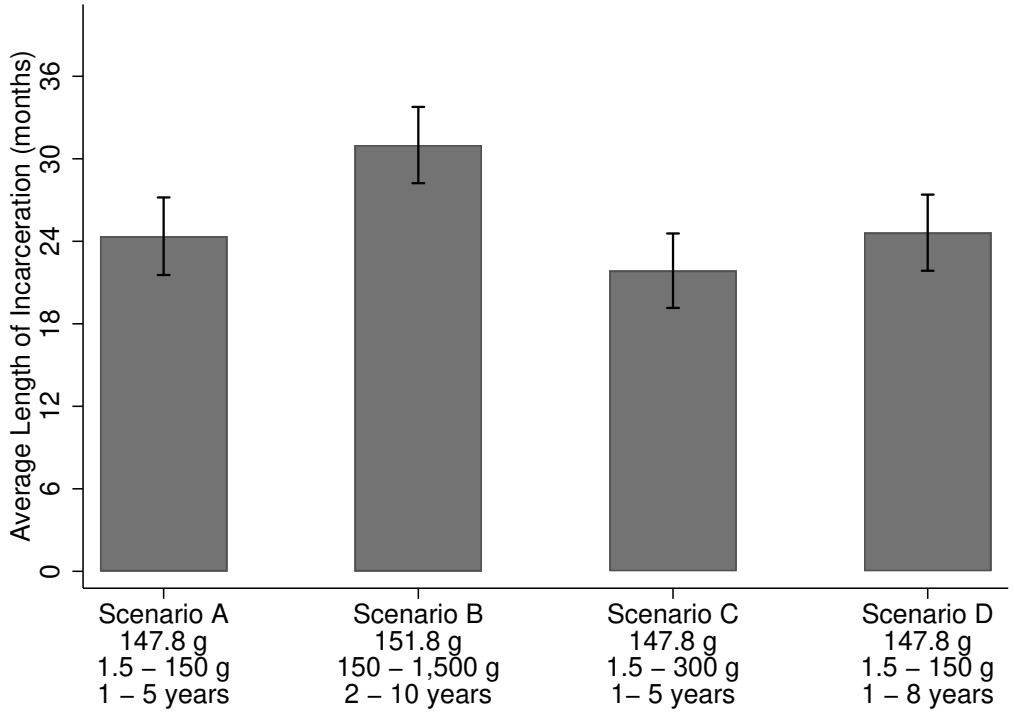
Figure 4 shows the average sentence recommended by prosecutors in different scenarios.[18] In scenario A, the average length of incarceration is slightly more than 2 years and in scenario B longer by more than 6 months. The figure also reveals that the lowest average sentence is in scenario C, slightly more than 22 months, and finally, scenario D is almost identical to scenario A.

Similarly to the theft case, Table 4 reports results from three specifications: (i) a simple univariate OLS regression – conveniently, in all three panels the treatment dummy equals zero for observations from scenario A; (ii) an OLS model controlling for additional characteristics of prosecutors (see Table 8 for full results); and (iii) a univariate OLS on a sample of district prosecutors. Panel I tests the overall threshold effect of 150g of methamphetamine, Panel II tests for the reference mechanism, and Panel III tests for the severity mechanism.

Panel I of Table 4 provides convincing evidence that the 150g 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 5.3 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%)) .

[18] Figure 8 in Appendix B shows the distribution of recommended sentences for each scenario.

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.

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***	5.342**
	(1.969)	(2.229)	(2.306)
Constant	24.370***	41.876	25.368***
	(1.407)	(25.440)	(1.778)
Control Variables	No	Yes	No
Prosecutors	All	All	District Off.
N	103	98	76

Panel II: Reference Effect

	Model 1	Model 2	Model 3
Treatment Effect	-2.506	-2.573	-5.448**
	(1.944)	(1.974)	(2.319)
Constant	25.379***	42.851	25.368***
	(1.407)	(26.304)	(1.782)
Control Variables	No	Yes	No
Prosecutors	All	All	District Off.
N	98	96	63

Panel III: Severity Effect

	Model 1	Model 2	Model 3
Treatment Effect	0.257	-0.808	-0.812
	(1.969)	(2.095)	(2.469)
Constant	24.370***	40.990***	25.368***
	(1.407)	(8.204)	(1.778)
Control Variables	No	Yes	No
Prosecutors	All	All	District Off.
N	105	102	74

Robust standard errors in parentheses

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

Notes: Panel I tests the overall threshold effect (s(B) - s(A)), Panel II tests the reference effect (s(C) - s(A)), and Panel III tests the severity effect (s(D) - s(A)). *Treatment* corresponds to a dummy variable which equals 0 for respondents who were assigned to the A scenario and 1 otherwise. Model 1 represents a simple univariate OLS regression, Model 2 extends the univariate OLS by controlling for additional characteristics of prosecutors. For full results see Table 8 in Appendix B. Model 3 is a univariate OLS on a sample of district prosecutors. Prosecutors who failed our attention check are dropped in all three specifications.

Panel II of Table 4 shows suggestive evidence of the reference mechanism. Models 1 and 2 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 months. The effect represents approximately a 10% decrease in the length of incarceration, but it is statistically insignificant. The reference mechanism decreases the average recommended sentence among district prosecutors by 5.4 months, which corresponds to a 20% decrease. The effect among district prosecutors is statistically significant. Interestingly, when we allow for gender specific treatment effect, the reference effect is also significant among male prosecutors. Table 8 reports the effect. Since the effect is statistically significant only on restricted samples, we consider it suggestive evidence of the reference effect.

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.[19] 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 from the perspective of proportionality. 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 on what is considered an unreasonably harsher sentence, we introduce a parametric measure.

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

[19]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.

consider a case (x, t) with a corresponding sentence s^A. Given this sentence s^A, we ask whether a different sentence s^B 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^B is *just* if it is: (i) at least as harsh as s^A; 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^B 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:

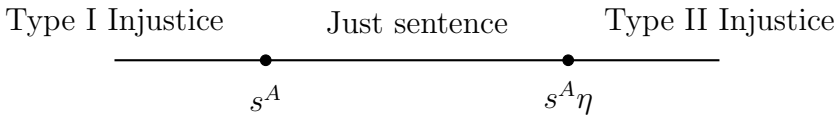
$$s^A \leq s^B \tag{D 2.1}$$

$$s^B \leq s^A \eta. \tag{D 2.2}$$

The tolerance parameter η captures what is considered a reasonably harsher sentence and what is not. Importantly, the definition says that a sentence s^B can be called *just* only if there is a tolerance parameter η and a sentence s^A to which s^B 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^B 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^A, then depending on the position of s^B, 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) and quantify the frequency of $\frac{s_j^B}{s_i^A}$ 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 pair of randomly observed sentences s_i^A and s_j^B, the latter is 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 150g 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 half of the sentences as *just*.

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

	$\mathcal{M}^I$	$\mathcal{M}^J(1.5)$	$\mathcal{M}^J(2)$	$\left(\mathcal{M}^J\right)^{-1}(0.5)$
Drug Possession: 150g threshold	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
Theft: Size of Damage	0.15	0.41	0.65	1.70

Notes: This table shows measures of Justice ($\mathcal{M}^J$) for two parameters of η and a measure of Type I injustice ($\mathcal{M}^J$) for four different comparisons. The last column shows what the η must be (how tolerant the society must be) to reach the level of 50% of just sentences.

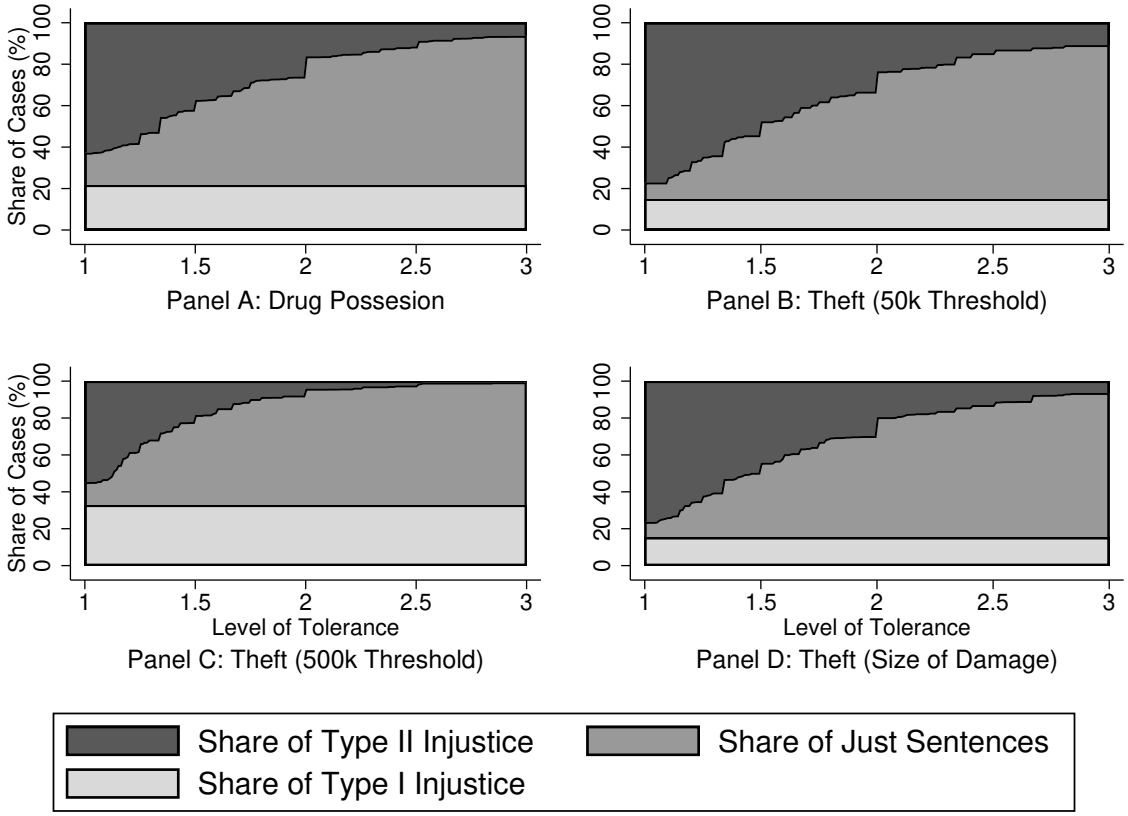
Table 5 and Figure 6 show the results for standardized parameters and graphically for $\eta \in (1, 3)$. The *Type I injustice* is highest in the 500k threshold comparison. In a third of all comparisons, the more severe cases were sentenced to a more lenient sentence. This suggests that these two cases (theft cases with damage of CZK 487,092 and CZK 508,213) are perceived the most similar among all four pairwise comparisons.

The extent of Type I injustice is driven by a large variation in recommended sentences by individual prosecutors documented in Figures 7 and 8 in Appendix B. Strikingly, in the majority of scenarios, the recommended sentences are close to both the lower and the upper limits of sentencing ranges. Quantity thresholds are thus not the only reason *just sentences* are not imposed.

The second column implies that if the society tolerates that a slightly 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 but also on the distribution of sentences.

Figure 6: Measures of (In)Justice in Experiment



Notes: Panel A shows measures of justice and injustice for the comparison of drug possession case scenarios A and B. Panel B and C show measures of justice and injustice for the case of theft around the 50k threshold and the 500k threshold. Panel D shows measures of justice and injustice for theft in terms of size of the damages. For any given parameter η between 1 and 3, the figure shows shares of comparisons considered Just sentences, Type I injustice, and Type II injustice.

5 Discussion

Introducing a measure providing guidance to sentencers necessarily includes a trade-off. Measures envisioned to reduce sentencing disparities will likely introduce a certain amount of new ones. As we show, the extent of new disparities may be substantial and the disparities introduced by dividing offenses into subsections with specific sentencing ranges may strongly limit the benefits of such a measure. The difference in sentences imposed by professional sentencers 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 quantity thresholds influence sentencing around them and what consequences result from changes in the sentencing range and offense subsection composition. 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 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 attenuates 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 the upper half of the sentencing range ($G(.) > 0.5$) are more sensitive to increases in the upper limit, while sentences in lower half of the sentencing range ($G(.) < 0.5$) are sensitive 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. Similarly, no life sentence has been imposed in Slovenia since its adoption in 2008, because judges consider it too severe a sanction (Filipčič, 2019).

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*.

An experimental design seems the most appropriate for future research into nuanced roles of thresholds' effects since defendants, police, and prosecutors respond to quantity thresholds (Bjerk, 2005, 2017b; Lepage, 2020; Travova, 2019), rendering any real-data sentencing study complicated. The cases studied need to be specifically designed so that it is difficult for defendants and state representatives to tailor their behavior to quantity thresholds. This 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 sentencing 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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Leibovitch, A. (2016a). Punishing on a curve. *Nw. UL Rev.*, 111:1205.

Leibovitch, A. (2016b). Relative judgments. *The Journal of Legal Studies*, 45(2):281–330.

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

Pease, K. and Sampson, M. (1977). Doing Time and Marking Time. *The Howard Journal of Criminal Justice*, 16(2):59–64.

Pintoff, R. (2004). The impact of incarceration on juvenile crime: a regression discontinuity approach. In *Technical Report, Working Paper*.

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

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

Skugarevskiy, D. (2017). *Essays in law and economics of enforcement*. Dissertation thesis, Graduate institute of international and development studies in Geneva.

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

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

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

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

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

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

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

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

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

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

Appendix A

Proof of Theorem 1

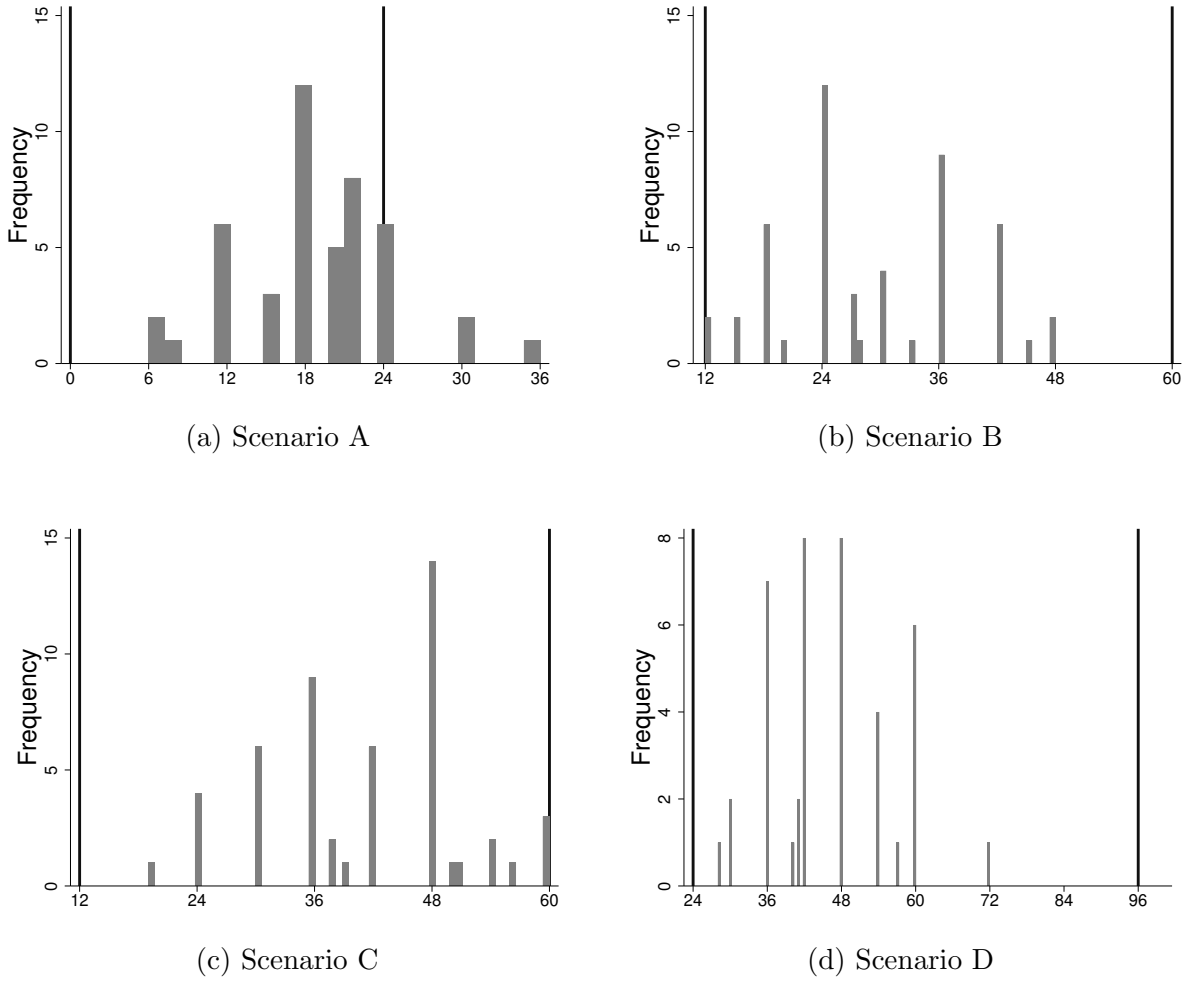
We show that equation T 1.1 holds. To simplify notation, label two cases A and B and define three differences $\Delta G = G(B) - G(A)$, $\Delta\rho^- = \rho^-(B) - \rho^-(A)$, and $\Delta\rho^+ = \rho^+(B) - \rho^+(A)$. Then, according to Definition 1, the difference in sentences $\Delta s = s(B) - s(A)$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

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

□

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. Note that prosecutors tend to round the length to years, which is in line with previous research; see Dhami et al. (2020); Pease and Sampson (1977). 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. Black vertical lines mark the upper and lower limits of the corresponding sentencing range. See Table 1 for scenario descriptions.

Table 6: Descriptive Statistics

Panel A: Theft

	A	B	C	D	balance test	not in exper.	H_0
Number of Observations	46	51	51	46		1049	
Male	43%	27%	51%	54%	0.067	55%	0.001
Communist Party	20%	16%	6%	13%	0.353	14%	0.841
Age	48	46.6	47	46.2	0.836	49.1	0.002
Time from Exam (years)	19.8	19	18.4	18.5	0.960	20.4	0.066
Time from Oath (years)	16.7	18.9	15.9	15.2	0.860	17.1	0.071
Alma Mater							
Brno	37%	37%	37%	35%	0.880	34%	0.529
Praha	39%	41%	37%	48%	0.497	42%	0.662
Plzen	7%	14%	10%	4%	0.401	9%	0.799
Olomouc	9%	0%	8%	4%	0.217	7%	0.840
Others	8%	8%	9%	10%	0.982	8%	0.762
Pros. General Office	11%	14%	4%	9%	0.308	4%	0.012
High Pros. Office	4%	2%	4%	9%	0.505	8%	0.147
Regional Pros. Office	15%	14%	20%	20%	0.819	24%	0.025
District Pros. Office	70%	71%	73%	63%	0.882	64%	0.114

Panel B: Drug Possession

	A	B	C	D	balance test
Number of Observations	58	50	44	54	
Male	41%	42%	34%	52%	0.361
Communist Party	10%	18%	10%	15%	0.537
Age	46.3	47.8	45.2	47.3	0.660
Time from Exam (years)	18.5	19	16.6	20.3	0.426
Time from Oath (years)	16.2	16.6	13.8	16.4	0.334
Alma Mater					
Brno	38%	38%	32%	38%	0.804
Praha	34%	40%	43%	46%	0.574
Plzen	16%	8%	11%	2%	0.087
Olomouc	5%	4%	7%	4%	0.895
Others	7%	10%	7%	11%	0.821
Pros. General Office	10%	6%	9%	9%	0.880
High Pros. Office	3%	2%	14%	2%	0.023
Regional Pros. Office	16%	14%	20%	17%	0.860
District Pros. Office	70%	78%	57%	72%	0.153

Notes: The H_0 column reports the p-value of the two-sided t-test under the null that the sample value of participants equals those who do not participate. The balance test represents the p-value of the F-test under the null that there is no difference between the groups.

Table 7: Theft: Full Results

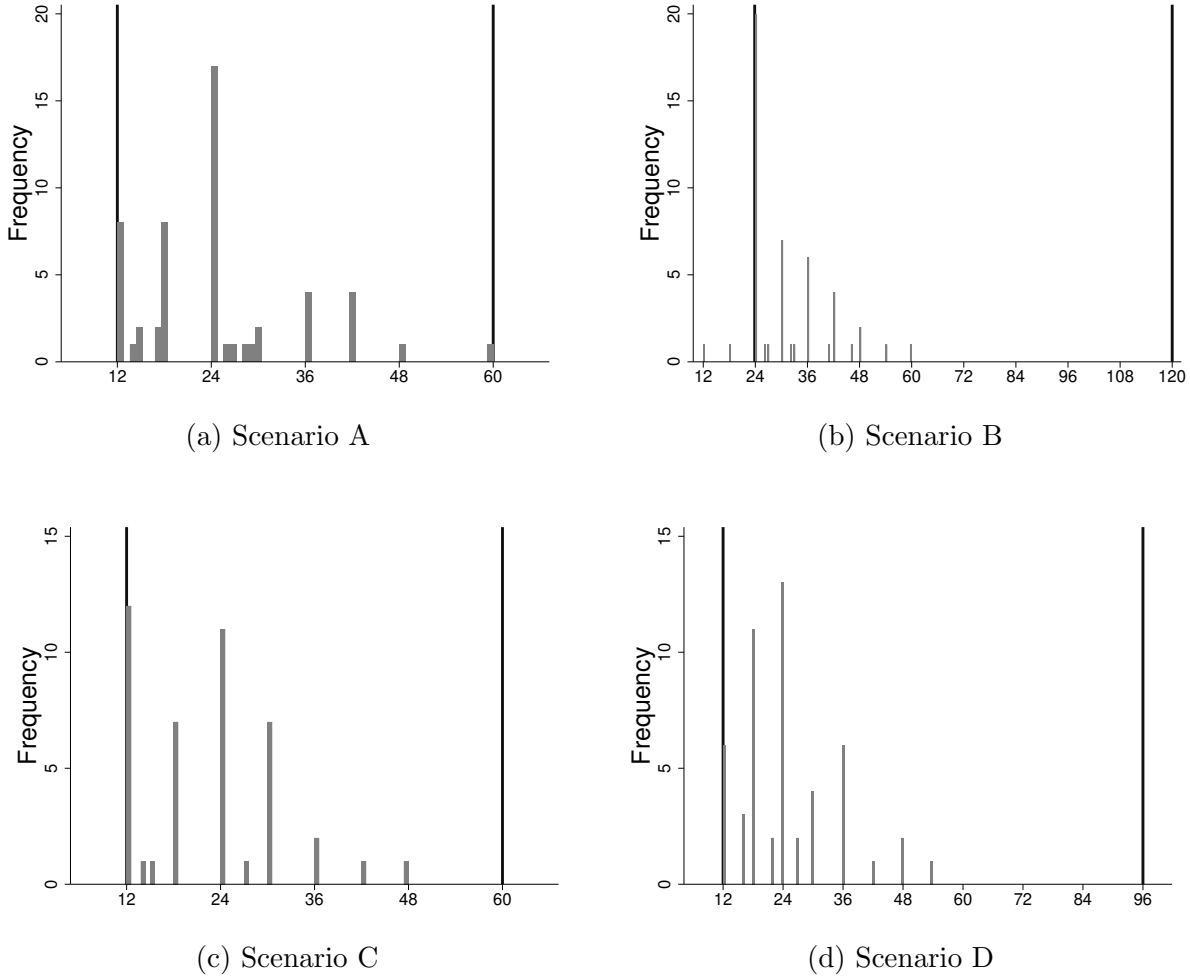
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
50k Threshold	11.57***	11.06***	14.75***						
	(1.7367)	(2.0596)	(3.2436)						
50k Threshold x Male		1.373							
		(4.0453)							
50k Threshold x District			-4.705						
			(3.8229)						
500k Threshold				4.676*	10.01***	2.388			
				(2.4299)	(3.7599)	(4.3265)			
500k Threshold x Male					-9.481**				
					(4.6727)				
500k Threshold x District						3.643			
						(5.6518)			
Size of Damage							11.65***	13.16***	13.01***
							(2.2633)	(2.6397)	(4.6103)
Size of Damage x Male								-4.079	
								(4.6758)	
Size of Damage x District									-1.883
									(5.3522)
Male	6.962***	6.301***	6.187***	-2.252	2.085	-2.645	4.079*	6.330*	4.107*
	(1.8738)	(2.0580)	(1.9473)	(2.2037)	(3.0929)	(2.4823)	(2.2290)	(3.3623)	(2.2644)
Age	0.134	0.142	0.106	-0.0684	-0.138	-0.0482	-0.0620	-0.00520	-0.0628
	(0.2337)	(0.2306)	(0.2463)	(0.2165)	(0.2138)	(0.2229)	(0.2574)	(0.2592)	(0.2550)
Alma Mater Brno	-7.343	-7.412	-6.573	4.336	7.148*	3.793	-1.152	-2.076	-1.094
	(5.1408)	(5.1953)	(4.7744)	(3.5502)	(3.6352)	(3.7564)	(6.4171)	(6.4281)	(6.5813)
Alma Mater Praha	-9.321*	-9.293*	-8.537*	5.094	7.837**	4.202	-1.310	-2.149	-1.401
	(5.2158)	(5.2971)	(4.8543)	(3.1275)	(3.1367)	(3.5974)	(6.3219)	(6.2592)	(6.4487)
Alma Mater Plzen	-8.459	-8.365	-7.474	6.995	10.59**	5.953	1.645	0.747	1.436
	(5.7492)	(5.8325)	(5.4676)	(4.7636)	(4.5703)	(5.3151)	(6.5335)	(6.5797)	(6.6516)
Alma Mater Olomouc	-7.169	-7.338	-5.896	4.084	7.453**	3.010	0.926	-0.392	0.655
	(5.1155)	(5.1419)	(4.7557)	(3.5068)	(3.5961)	(4.3165)	(7.2312)	(7.2354)	(7.2478)
Communist Party	-2.407	-2.532	-1.513	-3.682	-3.622	-4.423	1.516	0.459	1.465
	(3.7010)	(3.6225)	(3.9854)	(4.5892)	(4.6565)	(4.8029)	(5.1693)	(5.1689)	(5.1372)
Time from Oath	-0.0644	-0.0723	-0.0501	0.342	0.376	0.331	0.133	0.100	0.125
	(0.2362)	(0.2364)	(0.2416)	(0.2834)	(0.2748)	(0.2937)	(0.2739)	(0.2759)	(0.2814)
Regional Pros. Office	1.153	1.144	0.922	5.930	4.150	5.572	-0.00854	0.00576	0.0106
	(5.8133)	(5.8444)	(5.8032)	(6.0924)	(5.5660)	(5.9189)	(7.2147)	(7.2641)	(7.0376)
Pros. General Office	4.215	4.330	3.703	1.189	-1.266	0.933	3.933	3.857	4.310
	(5.7822)	(5.7860)	(5.6403)	(7.5144)	(6.9797)	(7.6957)	(7.3521)	(7.3780)	(7.2646)
District Pros. Office	-0.368	-0.202	1.935	-0.780	-1.339	-2.779	-4.686	-4.804	-3.658
	(5.2895)	(5.2853)	(5.8505)	(5.8798)	(5.1310)	(6.6351)	(6.6390)	(6.6905)	(6.7462)
Constant	18.16*	18.10*	17.11*	34.88***	33.31***	36.63***	31.91***	30.30***	31.35***
	(9.4668)	(9.4826)	(9.5244)	(10.0543)	(9.2715)	(10.2986)	(11.1759)	(11.1090)	(11.4204)
N	91	91	91	85	85	85	96	96	96

Standard errors in parentheses

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

Notes: The outcome variable is the recommended length of incarceration in months. The reference categories are Alma Mater Others and High Pros. Office.

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. Note that prosecutors tend to round the length to years, which is in line with previous research; see Dhami et al. (2020); Pease and Sampson (1977). 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. Black vertical lines mark the upper and lower limits of the corresponding sentencing range. See Table 3 for scenario descriptions.

Table 8: Drug Possession: Full Results

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
150g Threshold	6.005***	8.797***	10.33**						
	(2.1641)	(2.8175)	(4.2180)						
150g Threshold x Male		-6.130							
		(4.3047)							
150g Threshold x District			-5.905						
			(5.1130)						
Reference Effect				-1.879	1.741	4.296			
				(2.1236)	(2.3988)	(3.5105)			
Reference Effect x Male					-10.15**				
					(4.5792)				
Reference Effect x District						-10.06**			
						(4.2617)			
Severity Effect							-0.578	1.823	2.154
							(2.1522)	(2.4195)	(3.5253)
Severity Effect x Male								-5.100	
								(4.6052)	
Severity Effect x District									-3.921
									(4.1475)
Male	-0.278	2.469	-0.617	-1.078	3.135	-1.600	1.335	3.622	1.210
	(2.1987)	(3.2010)	(2.2445)	(2.1938)	(3.2868)	(2.2434)	(2.1937)	(3.2363)	(2.2362)
Age	0.181	0.156	0.107	0.00839	0.0743	0.0153	0.214	0.212	0.187
	(0.3398)	(0.3340)	(0.3122)	(0.1994)	(0.1984)	(0.1830)	(0.2504)	(0.2490)	(0.2512)
Alma Mater Brno	-10.27	-11.81	-10.76*	-10.84**	-11.74**	-12.12***	-10.42**	-11.91**	-10.73**
	(7.0379)	(7.2645)	(6.3828)	(5.2845)	(5.7788)	(4.3072)	(4.7407)	(4.9779)	(4.4242)
Alma Mater Praha	-9.188	-10.78	-9.644	-10.51**	-10.60*	-10.71**	-9.012*	-9.945**	-9.492**
	(7.3437)	(7.4324)	(6.6770)	(5.0576)	(5.5839)	(4.1939)	(4.5346)	(4.4548)	(4.1568)
Alma Mater Plzen	-9.727	-10.79	-10.35	-8.674	-8.025	-9.903**	-8.785	-9.877	-9.510*
	(7.4438)	(7.7426)	(6.7795)	(5.6027)	(6.2100)	(4.6369)	(5.7243)	(5.9571)	(5.5037)
Alma Mater Olomouc	-11.35	-12.51*	-10.72*	-9.212*	-11.27**	-10.40**	-12.04***	-13.16***	-12.38***
	(6.8594)	(7.1348)	(6.2226)	(5.0749)	(5.6115)	(4.1224)	(4.3209)	(4.3347)	(3.8073)
Communist Party	-0.0333	0.200	0.937	1.944	0.599	1.623	-2.502	-2.454	-2.081
	(4.1735)	(4.1670)	(4.1776)	(4.8792)	(4.8186)	(4.9411)	(3.7606)	(3.8351)	(3.8011)
Time from Oath	-0.126	-0.0745	-0.0282	0.114	0.0666	0.0934	-0.269	-0.277	-0.247
	(0.3993)	(0.3911)	(0.3730)	(0.2653)	(0.2554)	(0.2480)	(0.3320)	(0.3344)	(0.3353)
Regional Pros. Office	10.63***	9.362***	9.875***	5.854	5.469	7.472*	4.640	3.762	3.943
	(3.1273)	(3.4160)	(2.8957)	(4.4281)	(4.2231)	(3.9693)	(6.9132)	(6.9978)	(6.4677)
Pros. General Office	8.911*	7.380	8.432*	1.531	2.503	3.906	3.981	2.433	3.342
	(4.9846)	(5.0666)	(4.5511)	(5.0313)	(4.7297)	(4.6213)	(6.7978)	(6.9428)	(6.3223)
District Pros. Office	10.20***	8.765***	12.19***	3.404	2.666	9.598**	4.769	3.593	6.082
	(2.3714)	(2.7388)	(2.5615)	(3.8695)	(3.6879)	(4.2634)	(6.4937)	(6.6001)	(5.9097)
Constant	17.83	19.73	18.89	28.79***	25.74***	25.06***	23.42**	24.99**	24.03**
	(13.0217)	(13.0152)	(11.9782)	(8.7169)	(8.9020)	(7.9429)	(9.9865)	(9.8375)	(9.6065)
N	97	97	97	95	95	95	99	99	99

Standard errors in parentheses

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

Notes: The outcome variable is the recommended length of incarceration in months. The reference categories are Alma Mater Others and High Pros. Office.

Appendix C

Introduction Screen[20]

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.

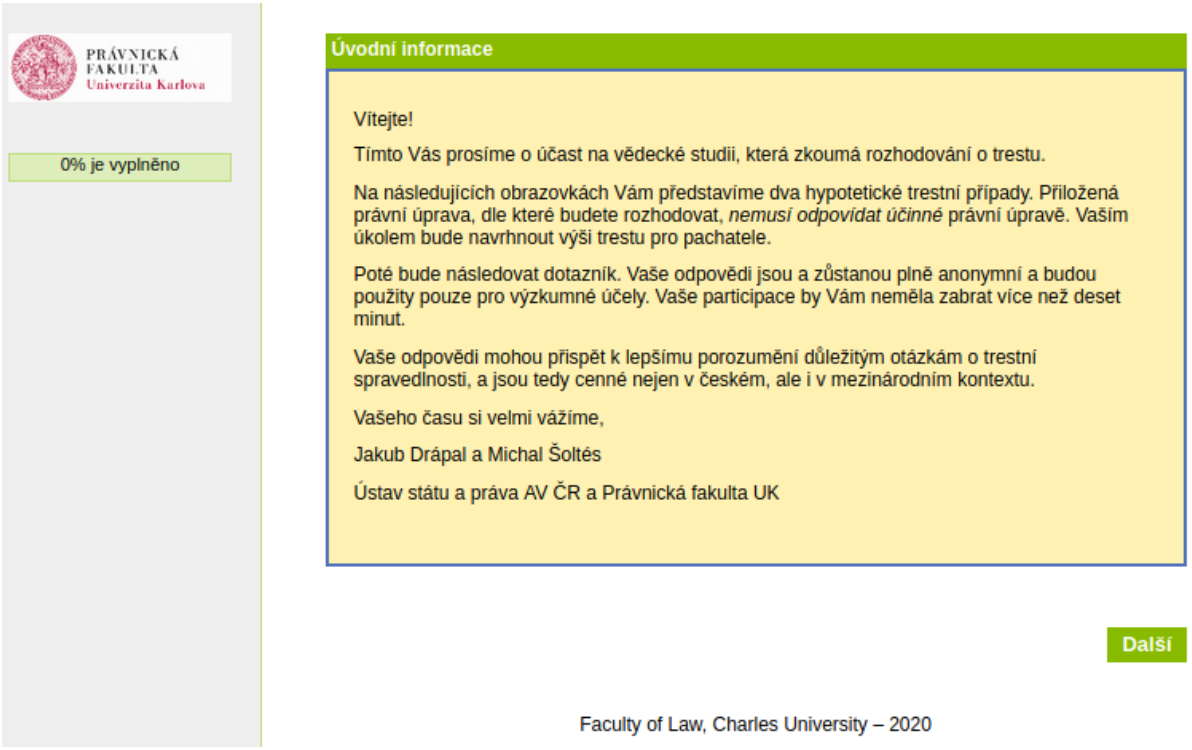
Your answers might contribute to the better understanding of important criminal justice issues that are significant not only for the Czech Republic, but for the international audience as well.

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

[20]Figure 9 shows screenshot.

Figure 9: Vignettes: Introduction Screen



Theft: Vignette[21]

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.

[21] Figure 10 shows screenshot.

Offense Section: Theft

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

...

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

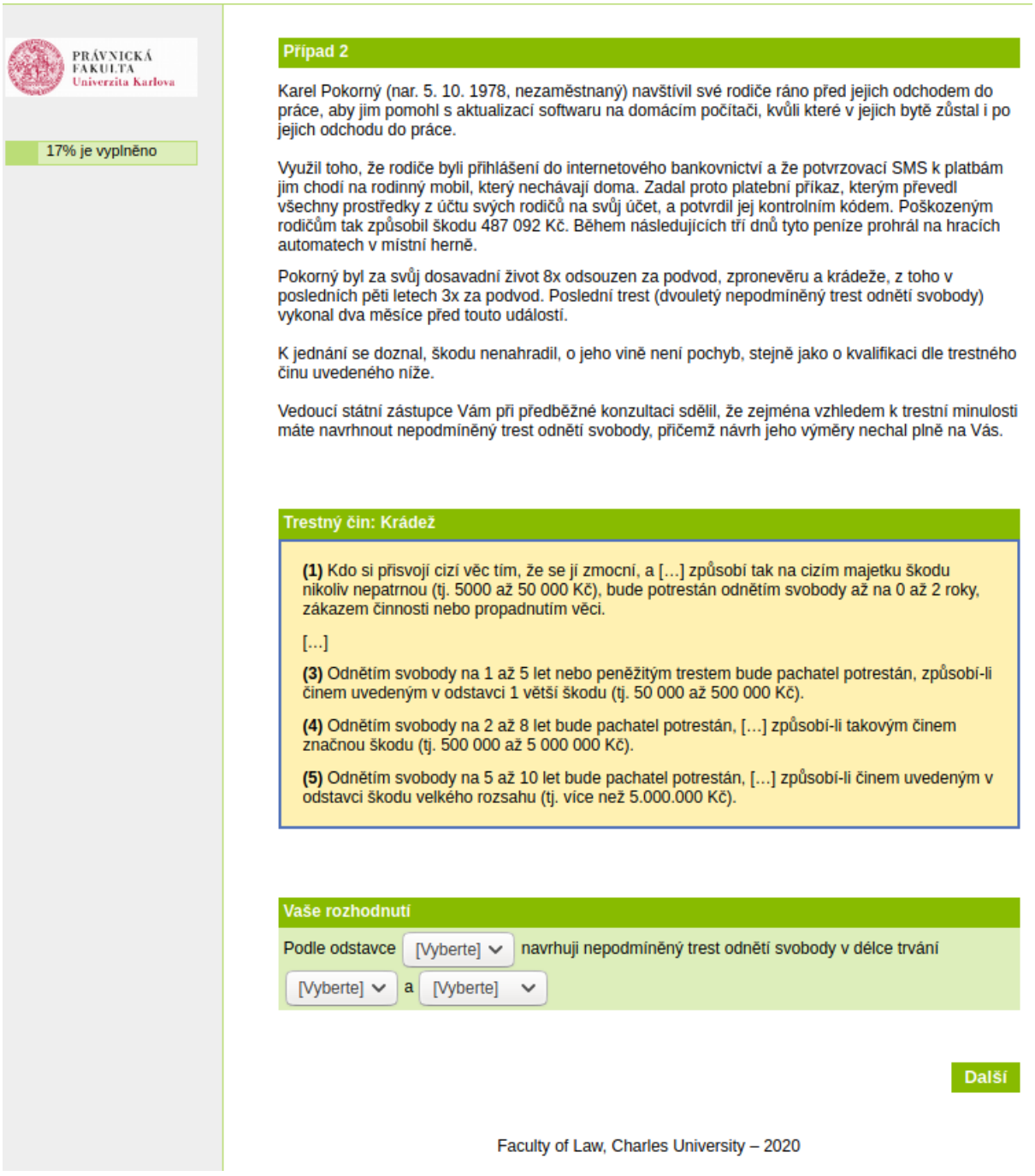
(4) An offender shall be sentenced to imprisonment for two to eight years, if he/she [...] causes substantial damage (meaning CZK 500,000-5,000,000) by such an act.

(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).

CASE DESCRIPTION

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: Theft: Vignette



Drug distribution: Vignette[22]

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.

[22]Figure 11 shows screenshot.

CASE DESCRIPTION

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

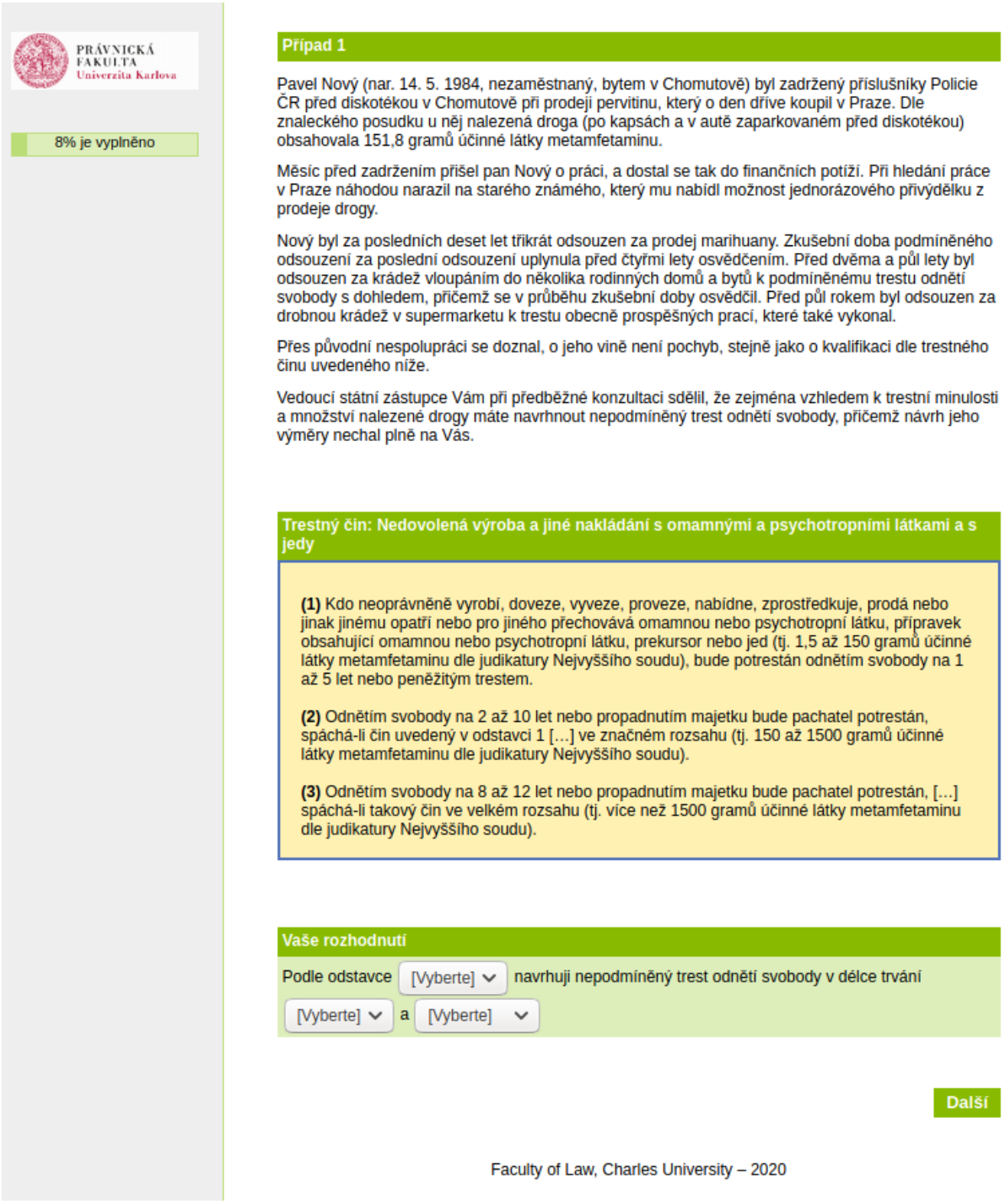
(1) Whoever produces, imports, exports, transports, offers, provides or sells or otherwise arranges for another or handles for another narcotic or psychotropic substances, products containing narcotic or psychotropic substances, precursors or poisons (meaning 1.5-150 grams of pure substance of methamphetamine 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.

(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 11: Drug Distribution: Vignette



Abstrakt

Trestní zákoník stanoví trestní sazby za jednotlivé trestné činy, čímž rámcově určuje, jaké tresty by měly být uloženy za typově vymezené trestné činy. S cílem přesněji stanovit typově vymezenou závažnost jsou běžně stanoveny odlišné trestní sazby za základní a kvalifikované či privilegované skutkové podstaty (odstavce). K tomuto odlišení jsou zákonem či judikaturou často použity kvantifikovatelné proměnné, jako je výše škody či množství účinné látky v případě drogových trestných činů. V článku zkoumáme dopady kvantifikovatelných hranic na uložené tresty. Tvrdíme, že vliv těchto hranic lze rozložit do dvou mechanismů: mechanismu závažnosti a referenčního mechanismu. Na základě experimentu se státními zástupci ukazujeme, že kvantifikovatelné hranice vedou ke skokovému zvyšování trestů, a tedy i k významným rozdílům při ukládání trestů ve velmi podobných případech. Dále definujeme nový empirický způsob měření (ne)spravedlnosti, který používáme pro vyčíslení dopadů kvantifikovatelných hranic na pravděpodobnost uložení spravedlivého trestu.

Working Paper Series
ISSN 1211-3298
Registration No. (Ministry of Culture): E 19443

Individual researchers, as well as the on-line and printed versions of the CERGE-EI Working Papers (including their dissemination) were supported from institutional support RVO 67985998 from Economics Institute of the CAS, v. v. i.

Specific research support and/or other grants the researchers/publications benefited from are acknowledged at the beginning of the Paper.

Published by
Charles University, Center for Economic Research and Graduate Education (CERGE)
and
Economics Institute of the CAS, v. v. i. (EI)
CERGE-EI, Politických vězňů 7, 111 21 Prague 1, tel.: +420 224 005 153, Czech Republic.
Printed by CERGE-EI, Prague
Subscription: CERGE-EI homepage: http://www.cerge-ei.cz

Phone: + 420 224 005 153
Email: office@cerge-ei.cz
Web: http://www.cerge-ei.cz

Editor: Byeongju Jeong

The paper is available online at http://www.cerge-ei.cz/publications/working_papers/.

ISBN 978-80-7343-522-6 (Univerzita Karlova, Centrum pro ekonomický výzkum a doktorské studium)
ISBN 978-80-7344-617-8 (Národohospodářský ústav AV ČR, v. v. i.)