Law Students Aren't Prosecutors, Yet: Evidence From Two Experiments*

Jakub Drápal†, Otakar Kořínek‡, and Michal Šoltés§

Preliminary Draft: Comments are Welcome

The most recent version

August, 2021

Abstract

Using experimental setting, we show how prosecutors and law students differ in their sentencing decisions. In line with existing literature, law students recommend harsher sentences than prosecutors in the same cases. Furthermore, we document that sentencing decisions by law students suffer from a higher degree of within-group sentencing inconsistency, limiting the possibilities of using students instead of professionals in experiments.

1 Introduction

Many experiments in social sciences rely on students as a primary participants (see e.g. Kantowitz et al., 2014). Using students has several practical benefits such as ease of a recruitment, but may limit the validity of results. A rich literature warns against generalizing students' results to general and/or to relevant population (Hanel and Vione, 2016; Peterson, 2001)[1]. A limited representativeness of students behaviour is likely even

*We thank Sven Hoeppner.

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

‡New PORG, Pod Krčským lesem 1300/25, Praha 4, 142 00 (Email: korinek@novyporg.cz)

§Department of Economics, Faculty of Law Charles University, nám. Curieových 901/7, 116 40 Prague, Czech Republic and CERGE-EI, a joint workplace of Charles University and the Economics Institute of the Czech Academy of Sciences, Politických vězňů 7, 111 21 Prague, Czech Republic. (Email: msoltes@prf.cuni.cz)

[1]Note that some scholars defend the use of student respondents (Glick, 2008)

more severe concern in activities and decisions that are usually made by well-educated and experienced professionals such as financial market traders, managers, judges and prosecutors (Cooper, 2006; Frechette, 2015). Criminal justice authorities are a prominent example of experts of such professionals.

Despite its limited generalizability, students have been commonly used in experimental sentencing literature. This includes papers studying effects of various defendant's characteristics such as gender, attractiveness, health, age, race/ethnicity (Mueller-Johnson and Dhami, 2009; Ahola et al., 2009, 2010; Leippe et al., 2017; Abwender and Hough, 2001), defendant's history of child abuse or neglect (Bell Holleran et al., 2016); behavioural biases such as the anchoring effect (Enough, Birte and Mussweiler, Thomas, 2001); a presence of criminal family members of defenders and victims Rerick et al. (2021); religiosity (Yelderman and Miller, 2017) on sentencing decisions.[2]

In this paper, we study differences in sentencing decisions between law students and prosecutors. We leverage data from experiments conducted with prosecutors (for more details see Drápal and Šoltés, 2021) and law students and compare their sentencing decisions in two criminal cases, each with four mildly differentiated scenarios. We ask two research questions. First, do law students and prosecutors differ in the average length of recommended sentence, i.e. in their level of harshness? Second, do law students and prosecutors differ in the within-group sentencing inconsistency?

The latter question has important methodological implications. Suppose law students and prosecutors differ only the their harshness level. In that case, research questions based on a difference in average recommended sentence between treatment and control groups could by answered by experiments with students. However, if distributions of recommended sentences differ in other aspects (i.e. sentencing inconsistency), then credible identification of sentencer responses using only student samples becomes more complicated. This is particular true if, due to higher sentencing inconsistency some respondents' responses are systematically more likely to be censored by a minimum and maximum extent of sanctions (e.g. possibility to impose suspended sentence only for sentences below a certain length of sentence). Furthermore, understanding differences between law students and prosecutors in recommended sentences beyond the difference in the average length of recommended sentence has a direct implication for using student samples to study sentencing inconsistency,

[2]Note that many of the cited studies increase external validity of their results by including additional subjects. Furthermore, in many of the cited studies, the findings are interpreted as evidence of behaviour of juries rather than professional sentences which is arguably more convincing. See for example Hosch et al. (2011).

2 Methods

2.1 Trial Design

We use data from vignette experiments with two distinct groups of participants: prosecutors and law students. Each participant in the experiment was asked to analyse and recommend a length of incarceration in two hypothetical criminal cases: (i) drug possession and (ii) theft. The cases were presented in this particular order to everyone. Each case was described on 1 page and the wording was consulted with several practising judges, so no relevant informant is omitted. We also provided the participants with the relevant offense section of the criminal code (with multiple subsections) together with the relevant section of the criminal code and excerpt from jurisprudence. The introductory screen warned the participants that the provisions might not be in line with those in force. Additionally, we asked the participants to indicate according to which subsection they sentence the offender. Since the classification of the subsection is based on quantifiable variables that were provided in the case, there is an objectively right answer. Failure to identify the correct subsections can be a sign of inattentive of careless response and the observations were not used.

There are four different scenarios for both criminal cases (drug possession and theft). The scenarios differ in the quantifying variable (amount of methamphetamine found and amount of money stolen, respectively), the sentencing range applicable, and/or the thresholds influencing the composition of cases within subsections. Notably, two of the drug possession case scenarios (1C and 1D) introduce legislation that differs from the existing one. Prosecutors thus cannot propose the same sentence they are trained in their real-life decisions. Furthermore, evidence from two criminal cases limits the possibility that the participants' sentencing decisions differ only in one type of crime. The only outcome we measure is the length of recommenced incarceration in months.

In the drug possession case, participants were asked to recommend a prison sentence for an offender who was selling methamphetamine in front of a dance club in a town in Northern Bohemia. It was a one-time event following the offender being fired from a job. He was repeatedly sentenced for selling marijuana over the last ten years and for small thefts. We have manipulated the amount of methamphetamine (its pure substance, henceforth only methamphetamine) found on him, the sentencing range applicable, and the thresholds influencing the composition of cases within subsections. The full vignette description, the corresponding criminal code section's text, and other information provided to the participants are presented in Appendix. The four scenarios that were applied are summarized in Panel A of Table 1.

Table 1: Parameters of the Individual Crime Scenarios

Panel A: Drug Possession Case			
Scenario	Amount Possessed (grams)	Subsection Composition (grams)	Sentencing Range (years of incarceration)
1A	147.8	1.5 – 150	1 – 5
1B	151.8	150 – 1500	2 – 10
1C	147.8	1.5 – 300	1 – 5
1D	147.8	1.5 – 150	1 – 8
Panel B: Theft Case			
Scenario	Damage Caused (CZK)	Subsection Composition (CZK)	Sentencing Range (years of incarceration)
2A	48 283	5 000 – 50 000	0 – 2
2B	51 762	50 000 – 500 000	1 – 5
2C	487 092	50 000 – 500 000	1 – 5
2D	508 213	500 000 – 5 000 000	2 – 8

The table describes the presented scenarios in the vignettes. Panel A shows the characteristics of the four scenarios used for the drug possession case while Panel B does so for the theft case. The panels show the quantifying variable (either amount of drugs possessed or the damage caused), the composition of crimes in the subsection the crime in each scenario falls into and the sentencing range for this subsection.

In the theft vignette, the participants were asked to recommend a prison sentence for an offender who stole from his parents. He visited his parents to help them with updating their computer and was left alone in their apartment. He took advantage of his parents being logged-in in their internet banking account and transferred all their money to his own account. He gambled away all of it in the following three days. He was previously sentenced for fraud and embezzlement. He was released from a two-year prison sentence two months before committing the offense in the vignette. The entire vignette with the criminal code section's text are presented in Appendix. Similarly to the previous case we randomized participants into four treatment arms (scenarios), as captured in Panel B of Table 1. All scenarios use existing legislation and vary only in the amount of magnifying variable, putting them into sentencing ranges composed of different cases.

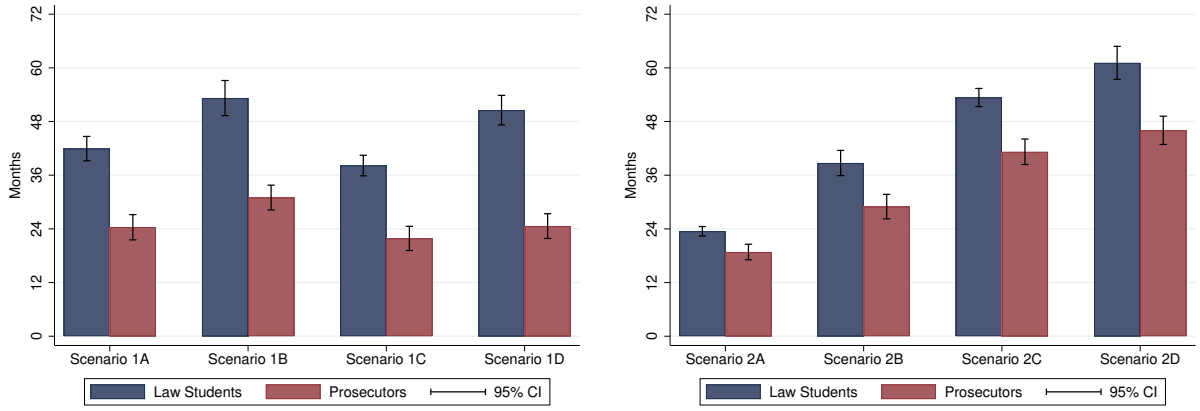
2.2 Participants

To approach prosecutors, we partnered with the Prosecutor General's Office of the Czech Republic. The invitation to participate in the survey was sent by a deputy to the Prosecutor General. 1257 prosecutors were 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 and match the experimental data with anonymized administrative data containing basic information about prosecutors. See Table 3 in Appendix for descriptive statistics.

We conducted the same experiment with law students from the Faculty of Law, Charles University. 752 students recommended a sentence in one of the four scenarios in the drug case and 572 in the theft case. We analyse decisions of 692 and 484 students who identified the right paragraph according to which they recommended the sentence in the drug possession and theft vignettes, respectively.[3] We collect students' recommended sentences, basic demographic characteristics, perception regarding the harshness of the criminal justice system and whether they passed the criminal law exam. Additionally, participants completed a questionnaire about their political preferences, age, and the purpose they think a punishment should have. See Table 4 in Appendix for descriptive statistics.

[3]Number of students in the drug possession and the theft cases differs because in the latter they were randomized into two additional scenarios which we do not study, as we do not have counterpart decisions from prosecutors.

Figure 1: Average Length of Recommended Sentence



(a) Panel 1

(b) Panel 2

Notes: Figure shows the average length of recommended sentences by law students (blue) and prosecutors (red). Panel 1 captures four scenarios of the drug possession case. 1A: 147.8g of methamphetamine under existing law; 1B: 151.8g of methamphetamine under existing law; 1C: 147.8g grams of methamphetamine and 300 grams as a threshold; 1D: 147,8g of methamphetamine and (1-8) years as sentencing range. Panel 2 captures four scenarios of the theft case. 2A: damage of CZK 48,283; 2B: damage of CZK 51,762; 2C: damage of CZK 487,092; 2D: damage of CZK 508,213

3 Results

We next provide empirical evidence that: (i) on average, prosecutors recommend more lenient sentences than law students; and (ii) prosecutors, as a group, tend to be less inconsistent in sentencing than law students.

3.1 Average Sentences

Our first finding is that students recommend substantially longer sentences in all compared scenarios. Panel 1 of Figure 1 shows average length of recommenced sentences in four scenarios of the drug possession case. In scenario D, students recommended twice as long incarceration as prosecutors did. Panel 2 captures average recommended length of sentences in the theft case. Similarly, students recommended a significantly longer sentence than prosecutors in all scenarios of the theft case.

We further show that students' characteristics, such their perceptions of harshness of the criminal justice system and whether they have passed criminal law exam, do not fully explain the difference between students and prosecutors. Figures 2 and 3 in Appendix capture recommended lengths of sentences by groups of students according to these characteristics. In all scenarios the general patterns are the same; prosecutors recommend the shortest incarceration. Interestingly, even students who perceive the current criminal justice system as strict recommend longer sentences than those recommended by prose-

cutors.[4] We thus argue that while students' harshness perception and preferences play a role in their sentencing decisions, even the most lenient students recommend sentences longer than prosecutors.

3.2 Sentencing Inconsistency

Within-group sentencing inconsistency among prosecutors tends to be substantially lower than among students. In six scenarios, sentences recommended by prosecutors have significantly lower standard deviations. In the two remaining scenarios, the standard deviations of prosecutors' and law students' sentences are statistically undifferentiated. Furthermore, we pairwise compare all sentences recommended by prosecutors (students) in a given scenario and quantify the average ratio of the compared sentences and the probability that the ratio of two randomly compared sentences exceed a given magnitude (e.g., one sentence is more than a double of the other one).

Panel 1 of Table 2 captures different measures of inconsistency among law students and prosecutors in the drug possession case. The third column shows the standard deviations of all sentences recommended by corresponding subjects. The standard deviations among the law students are substantially higher in all cases and often more than double. The fourth column shows p-values of the null hypothesis of equality of standard deviation. In all four cases, we can convincingly reject the equality. The fifth column shows the average ratio of all pairwise comparisons of recommended sentences. Finally, the last two columns show the probability that a ratio of two randomly compared sentences will exceed a magnitude of 2 and 3, respectively. For example, the probability from two randomly observed sentences recommended by law students in scenario 1A the longer will be more than double the shorter one is 25%, while among prosecutors it is only 17%. Overall, within-group sentencing inconsistency among prosecutors is lower than among students regardless of the applied measure.

[4]Note that some of the differences are not statistically significant.

Table 2: Inconsistency and Severity of Students' and Prosecutors' Recommended Sentences

	Obse.	SD	Equality of SD	Average ratio	Ratio > 2	Ratio > 3
Panel 1: Drug Possession Case						
Scenario 1A						
Law Students	168	17.9	0.000	1.8	25%	10%
Prosecutors	54	10.3		1.7	17%	4%
Scenario 1B						
Law Students	163	25.5	0.000	1.9	30%	10%
Prosecutors	49	9.7		1.4	7%	1%
Scenario 1C						
Law Students	178	15,5	0.000	1.8	22%	8%
Prosecutors	44	8.9		1.7	18%	3%
Scenario 1D						
Law Students	183	22.7	0.000	2.0	29%	12%
Prosecutors	51	9.8		1.6	15%	2%
Panel 2: Theft Case						
Scenario 2A						
Law Students	126	6	0.928	1.4	7%	3%
Prosecutors	46	5.9		1.6	12%	4%
Scenario 2B						
Law Students	123	15.8	0.000	1.8	25%	11%
Prosecutors	50	9.6		1.6	13%	2%
Scenario 2C						
Law Students	124	11.4	0.343	1.3	8%	0%
Prosecutors	51	10.1		1.4	5%	0%
Scenario 2D						
Law Students	117	20.1	0.000	1.6	15%	3%
Prosecutors	41	10.0		1.3	1%	0%

The table shows several measures of within-group inconsistency. Panel 1 compares situations between law students and prosecutors in the drug possession case, whereas Panel 2 in the theft case. Column Equality of SD presents p-value of a test with the null hypothesis that the standard deviation of law students and prosecutors are the same. The last two columns represent the probability that a ratio of two randomly compared sentences will exceed 2 and 3, respectively.

Panel 2 of Table 2 shows the same measures for the theft case. The main story remains the same. The standard deviation among prosecutors is lower in all four scenarios. However, the difference is statistically significant only in two scenarios. Our other measures of sentencing inconsistency provide additional evidence for higher within-group inconsistency among law students. The only exception is scenario 2A, in which the average ratio and so do the shares of extreme sentences are higher among prosecutors. Scenarios 2B and 2D again provide convincing evidence that sentencing inconsistency is lower among prosecutors.

4 Discussion

We study differences in sentencing decisions between law students and prosecutors and show that in the same criminal cases with the same information set, prosecutors recommend more lenient sentences than law students. We further provide novel evidence that the within-group sentencing inconsistency tends to be lower among prosecutors. Prosecutors as a group are thus more consist in their sentencing decision.

Our findings have two implications. The first one is a methodological one. Using law students (or potentially any other laypeople) as subjects in sentencing experiments carries more difficulties than just the different levels of harshness. Since law students tend to be more inconsistent, the distribution of recommended sentences generated by law students is not only a shifted distribution of professionals, but it differs in other properties. That further limits the validity of any results obtained using law students in sentencing experiments and especially in research of sentencing inconsistency.[5]

The second implication follows from the fact that prosecutors tend to be more consistent. If one accepts the premise that sentencing consistency is a desired feature of the criminal justice system, it means that prosecutors (professionals) do a better job than students would. A similar normative statement (prosecutors do better than students) could not be drawn from a simple comparison of averages of sentences.

References

D. A. Abwender and K. Hough. Interactive effects of characteristics of defendant and mock juror on us participants' judgment and sentencing recommendations. *The Journal of social psychology*, 141(5):603–615, 2001.

[5]Furthermore, our results have methodological implications for experimental design and power calculations. Since the standard deviation among prosecutors is lower, researchers would need fewer professionals than students to identify the same effect.

A. S. Ahola, S. Å. Christianson, and Å. Hellström. Justice needs a blindfold: Effects of gender and attractiveness on prison sentences and attributions of personal characteristics in a judicial process. *Psychiatry, Psychology and Law*, 16(sup1):S90–S100, 2009.

A. S. Ahola, Å. Hellström, and S. Å. Christianson. Is justice really blind? effects of crime descriptions, defendant gender and appearance, and legal practitioner gender on sentences and defendant evaluations in a mock trial. *Psychiatry, psychology and law*, 17(2):304–324, 2010.

L. L. Bell Holleran, T. J. Vaughan, and D. M. Vandiver. Juror decision-making in death penalty sentencing when presented with defendant's history of child abuse or neglect. *Behavioral sciences & the law*, 34(6):742–766, 2016.

D. J. Cooper. Are experienced managers experts at overcoming coordination failure? *Advances in Economic Analysis & Policy*, 5(2), 2006.

J. Drápal and M. Šoltés. Sentencing decisions around quantity thresholds: Theory and experiment. *Available at SSRN 3862637*, 2021.

Enough, Birte and Mussweiler, Thomas. Sentencing under uncertainty: Anchoring effects in the courtroom 1. *Journal of applied social psychology*, 31(7):1535–1551, 2001.

G. Frechette. Laboratory experiments: Professionals versus students. In *Handbook of Experimental Economic Methodology*, pages 360–390. Oxford University Press, 2015.

P. Glick. Restating the case: The benefits of diverse samples for theory development. *Psychological Inquiry*, 19:78–83, 04 2008. doi: 10.1080/10478400802049886.

P. Hanel and K. Vione. Do student samples provide an accurate estimate of the general public? *PLoS ONE*, 11, 12 2016. doi: 10.1371/journal.pone.0168354.

H. M. Hosch, S. E. Culhane, V. A. Tubb, and E. A. Granillo. Town vs. gown: A direct comparison of community residents and student mock jurors. *Behavioral Sciences & the Law*, 29(3):452–466, 2011.

B. H. Kantowitz, H. L. Roediger III, and D. G. Elmes. *Experimental Psychology*. Cengage Learning, 2014.

M. R. Leippe, A. N. Bergold, and D. Eisenstadt. Prejudice and terror management at trial: Effects of defendant race/ethnicity and mortality salience on mock-jurors' verdict judgments. *The Journal of social psychology*, 157(3):279–294, 2017.

K. U. Mueller-Johnson and M. K. Dhami. Effects of offenders' age and health on sentencing decisions. *The Journal of Social Psychology*, 150(1):77–97, 2009.

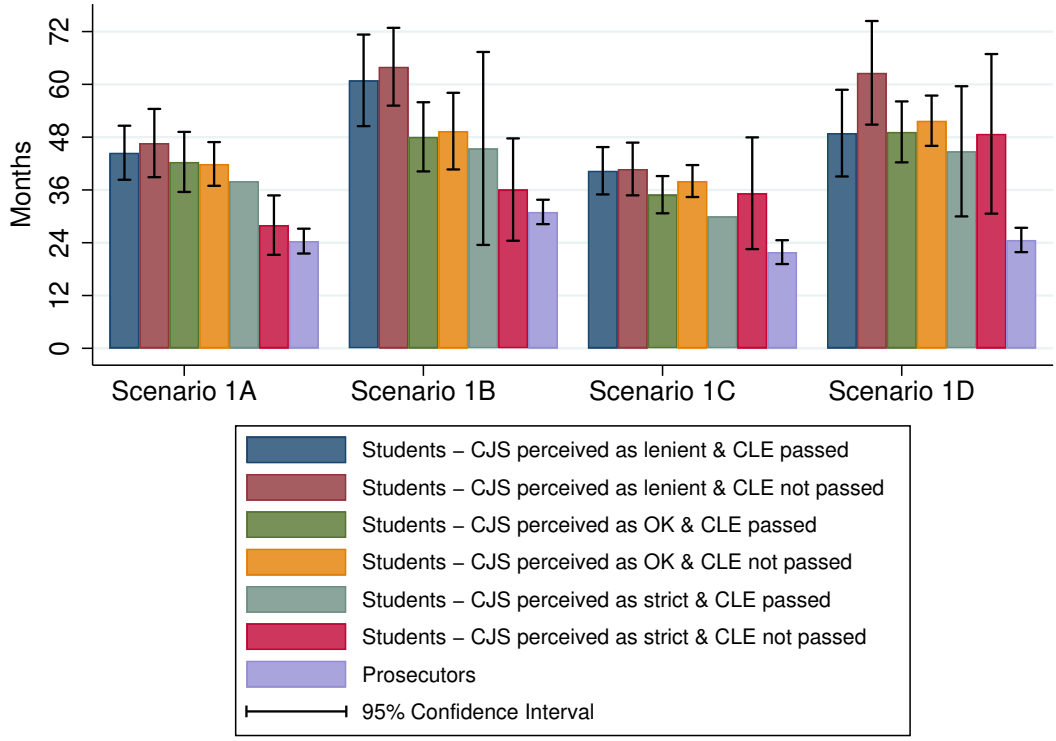
R. Peterson. On the use of college students in social science research: Insights from a second-order meta-analysis. *Journal of Consumer Research*, 28:450–61, 02 2001. doi: 10.1086/323732.

P. O. Rerick, T. N. Livingston, and M. K. Miller. Guilt by association: mock jurors' perceptions of defendants and victims with criminal family members. *Psychology, Crime & Law*, 27(3):282–305, 2021.

L. A. Yelderman and M. K. Miller. Religious fundamentalism, religiosity, and priming: Effects on attitudes, perceptions, and mock jurors' decisions in an insanity defense case. *Psychology, Crime & Law*, 23(2):147–170, 2017.

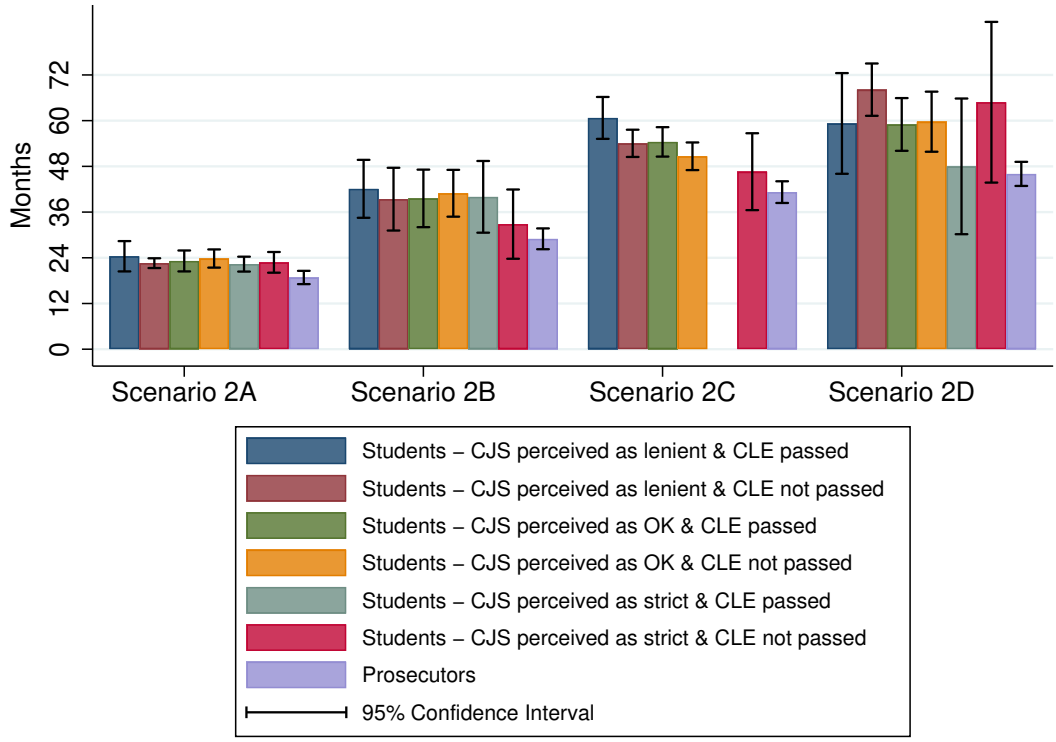
Appendix

Figure 2: Drug Possession: Average Recommended Sentences by Groups of Respondents



Notes: Figure shows the average recommended sentences in four scenarios in the drug possession cases imposed by law students and prosecutors. The group of students is split into six different categories according to their perception of harshness of the criminal justice system in the Czech Republic and whether they have passed exam from criminal law. Prosecutors tend to recommend the shortest length of incarceration. Students who perceive the criminal justice system as more strict tend recommended shorter sentences.

Figure 3: Theft: Average Recommended Sentences by Groups of Respondents



Notes: Figure shows the average recommended sentences in four scenarios in the theft cases imposed by law students and prosecutors. The group of students is split into six different categories according to their perception of harshness of the criminal justice system in the Czech Republic and whether they have passed exam from criminal law. Prosecutors tend to recommend the shortest length of incarceration.

Table 3: Descriptive Statistics of Prosecutors

Panel A: Drug Possession

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

Panel B: Theft

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

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

Table 4: Descriptive Statistics of Students

Panel A: Drug Possession

Scenario	1A	1B	1C	1D
Number of Observations	169	165	179	183
Male	48%	58%	49%	43%
Female	38%	33%	42%	44%
Age	22.8	23.0	23.5	23.0
Passed Exam	33%	36%	39%	36%
Perception of Czech Criminal Justice System				
Lenient	33%	36%	28%	25%
OK	43%	35%	50%	42%
Harsh	9%	11%	6%	11%

Panel B: Theft

Scenario	2A	2B	2C	2D
Number of Observations	126	123	124	117
Male	46 %	50 %	50 %	61
Female	52 %	41%	48 %	36
Age	23.1	23.0	23.0	23.2
Passed Exam	44%	38%	37%	46%
Perception of Czech Criminal Justice System				
Lenient	32 %	28 %	32 %	29 %
OK	44 %	39 %	52 %	53 %
Harsh	13 %	15 %	8 %	9 %

Notes: The table displays the demographic characteristics of the students (students who failed to identify the right paragraph were not included). Variable *passed exam* is a dummy whether students passed their criminal law exam. Perception of Czech criminal justice system measures whether student consider the CCJS lenient, harsh or appropriately strict. The values need not add to 100 %, as some have no opinion.

Table 5: Average Recommended Sentences

Panel A: Drug Possession Case

Scenario	1A	1B	1C	1D
Prosecutor	-17.576***	-22.239***	-16.294***	-25.914***
	(2.569)	(3.731)	(2.438)	(3.261)
Constant	41.946***	53.239***	38.157***	50.541***
	(1.267)	(1.794)	(1.085)	(1.522)

Standard errors in parentheses

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

Panel B: Theft Case

Scenario	2A	2B	2C	2D
Prosecutor	-4.642***	-9.744***	-12.136***	-15.095***
	(1.024)	(2.397)	(1.838)	(3.280)
Constant	23.468***	38.724***	53.371***	61.120***
	(0.530)	(1.239)	(0.992)	(1.671)

Standard errors in parentheses

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

Notes: Panels A and B show results from OLS regressions of the recommended sentence length in months. A dummy variable *Prosecutors* equals 1 for prosecutors and 0 otherwise.