The Conviction Lottery: Legal Indeterminacy and Judicial Discretion in Brazil's Drug Courts*

APEP Autonomous Research[†]   @SocialCatalystLab

March 31, 2026

Abstract

When criminal law gives judges broad discretion over high-stakes classifications, does random case assignment produce arbitrary punishment? We study this question in São Paulo's drug courts, where an electronic lottery assigns cases to 31 courtrooms. Using administrative records on 29,348 criminal cases across three offense types—drug trafficking, robbery, and theft—we document that drug trafficking conviction rates are essentially uncorrelated with theft conviction rates across the same courtrooms ($r = 0.10$), even though robbery and theft conviction rates are strongly correlated ($r = 0.67$). A common courtroom severity factor explains 83% of between-vara variance in robbery and theft but only 10% for drug trafficking. The remaining 90% is not explained by the common severity dimension shared by clearer offenses—consistent with an offense-specific discretion channel activated by Brazil's threshold-free drug statute. This *discretion decoupling* is left-skewed: some courtrooms are dramatically lenient on drugs while maintaining normal conviction rates for other offenses. The vague legal standard does not merely increase dispersion—it opens a separate dimension of judicial behavior that does not exist under clearer statutes.

---

1. Introduction

Modern legal systems face a fundamental design choice: should the law specify precise rules that constrain judicial decisions, or should it articulate broad standards that grant judges flexibility? The answer matters most where the stakes are highest. In criminal law, the choice between a bright-line quantity threshold and a multifactor "totality of the circumstances" standard can determine whether a defendant walks free or serves five years in prison.

This paper studies what happens when the law chooses the standard over the rule. Brazil's drug statute, *Lei* 11.343/2006, imposes a minimum five-year sentence for trafficking but provides no objective criteria—no quantity thresholds, no weight cutoffs, no bright-line rules—to distinguish a trafficker from a user. The classification is left entirely to judicial discretion. We exploit the electronic lottery (*sorteio*) that randomly assigns cases to criminal courtrooms (*varas*) in São Paulo to measure the consequences of this indeterminacy.

Our main finding is not that drug trafficking conviction rates vary across courtrooms—judicial heterogeneity is well-documented (Kling, 2006; Dobbie et al., 2018). Rather, we show that drug trafficking conviction rates operate on a *separate judicial dimension* from other offenses handled by the same courtrooms. Within the 31 criminal varas of São Paulo's Central courthouse, robbery and theft conviction rates are strongly correlated ($r = 0.67$): courtrooms that are harsh on one property crime are harsh on the other. But drug trafficking conviction rates are essentially uncorrelated with theft ($r = 0.10$) and only moderately correlated with robbery ($r = 0.47$). A common courtroom severity factor—extracted from robbery and theft—explains 83% of between-vara variance for those offenses but only 10% for drug trafficking. The remaining 90% is offense-specific discretion.

We call this pattern *discretion decoupling*: the vague drug standard opens a dimension of judicial behavior that does not exist under clearer statutes. This decoupling is not symmetric. The trafficking-specific component is left-skewed: some courtrooms convict only 40–55% of drug trafficking defendants while maintaining 70–96% conviction rates for robbery. No courtroom displays the reverse pattern of extreme drug-specific harshness. The vague standard creates room for leniency—and some judges use it.

The institutional setting makes this finding unusually credible. São Paulo's *sorteio eletrônico* is a centralized electronic lottery mandated by the *Conselho Nacional de Justiça* (CNJ) that distributes cases across courtrooms within the same assignment pool. In Central, all 31 varas draw from the same pool of cases. We verify that the *sorteio* is the dominant assignment mechanism (99.6% coverage for trafficking) and that filing day-of-week—a pre-determined characteristic—is uncorrelated with vara leniency. A leave-one-out leniency instrument yields a first-stage coefficient of 0.97 ($t = 33.4$), indicating that vara assignment

is the dominant predictor of individual conviction outcomes.

The comparison across offense types is the paper's key contribution. Because the same courtrooms handle drug trafficking, robbery, and theft, and because the same lottery assigns all three offense types, any difference in the *structure* of conviction rates across offenses is unlikely to reflect courtroom composition or case routing, since all three offense types pass through the same lottery to the same courtrooms. While the underlying factual content of cases differs across offenses, the differential correlation structure is consistent with the hypothesis that the vague drug standard activates an offense-specific discretion channel that clearer statutes do not.

This paper contributes to three literatures. First, we contribute to the economics of judicial discretion and the judge-leniency design tradition (Kling, 2006; Maestas et al., 2013; Dobbie et al., 2018; Bhuller et al., 2020; Norris et al., 2024). These papers use random assignment to estimate the causal effects of judicial decisions on downstream outcomes. We use the same institutional variation for a different purpose: measuring the *dimensional structure* of judicial behavior across offenses with different levels of legal determinacy. This reveals that judge heterogeneity is not a single dimension—it has offense-specific components that are activated by vague legal standards.

Second, we contribute to the law-and-economics literature on rules versus standards (Ehrlich and Posner, 1974; Kaplow, 1992; Sunstein, 1995). That literature has theorized that standards grant decision-makers more flexibility but increase outcome variance. We provide the first empirical test of this prediction within a single institutional setting where cases are randomly assigned to the same decision-makers across offense types governed by standards of different precision. The comparison design holds the decision-maker constant, isolating the effect of legal indeterminacy on the structure of judicial outcomes.

Third, we contribute to the literature on drug policy and mass incarceration in Brazil. Brazil incarcerates 920,000 people—third globally after the United States and China—and 28% are drug offenders. The *Supremo Tribunal Federal* (STF) ruled in June 2024 that personal cannabis possession does not constitute a crime, establishing a 40-gram threshold for cannabis only (RE 635659)—but this narrow ruling leaves the broader classification problem for other substances unresolved. Our evidence that the current standard creates a separate dimension of judicial discretion, largely on the lenient side, speaks directly to this constitutional debate.

The paper proceeds as follows. Section 2 reviews the related literature. Section 3 describes the institutional background. Section 4 presents the data. Section 5 details the empirical strategy. Section 6 reports results. Section 7 discusses mechanisms and implications.

2. Related Literature

This paper sits at the intersection of three literatures: judicial discretion and random assignment designs, the economics of rules versus standards, and the political economy of criminal justice in developing countries.

2.1 Judicial Discretion and Leniency Designs

A growing body of work exploits the random assignment of cases to judges or examiners to study the consequences of judicial discretion. Kling (2006) pioneered the use of random judge assignment to estimate the effects of incarceration length on employment and earnings. Dobbie et al. (2018) use judge leniency in the bail decision to show that pretrial detention causally increases conviction rates and future crime. Aizer and Doyle (2015) demonstrate that juvenile incarceration reduces high school completion and increases adult incarceration. Bhuller et al. (2020) find that incarceration reduces recidivism in Norway, while Norris et al. (2024) find it increases recidivism in the United States, highlighting the role of prison conditions and reentry support.

These papers share a common structure: random assignment creates variation in judicial severity, which is then used as an instrument for the downstream treatment of interest (incarceration, detention, disability receipt). Our paper uses the same institutional variation but asks a different question. Rather than estimating the effect of a single judicial decision, we examine the *dimensional structure* of judicial behavior across offenses. This reveals that judge heterogeneity is not one-dimensional—it has offense-specific components that are activated by the legal standard governing each offense.

Hull (2025) has recently formalized the econometric challenges of judge-leniency designs, particularly the "bundle" problem: judges who differ in one dimension (conviction) typically differ in others (pretrial detention, case management), making exclusion restrictions difficult to defend. Our design sidesteps this concern because we study the *structure* of bundles across offenses rather than attempting to isolate any single causal channel.

2.2 Rules Versus Standards

The theoretical literature on legal form has long recognized that the choice between rules and standards involves trade-offs in administrative cost, predictability, and flexibility (Ehrlich and Posner, 1974; Kaplow, 1992). Rules are more expensive to create (the legislature must specify content ex ante) but cheaper to apply (the judge simply checks whether the rule's conditions are met). Standards are cheaper to create (a vague directive suffices) but more

expensive to apply (the judge must determine content ex post, case by case). The key prediction for our purposes is that standards increase decision-maker heterogeneity: because the standard's content is determined at the point of adjudication, different judges applying the same standard to similar cases may reach different conclusions (Kaplow, 1992).

Sunstein (1995) argues that this heterogeneity is sometimes a feature rather than a bug, particularly in domains where rules would be too rigid to accommodate legitimate variation. Korobkin (2000) adds a behavioral dimension, showing that cognitive biases among decision-makers can amplify the variance introduced by standards beyond what a rational-actor model would predict.

Our paper provides the first within-setting empirical test of the prediction that standards increase judicial heterogeneity. Prior empirical work has compared sentencing dispersion across jurisdictions with different legal regimes (Abrams et al., 2012), which confounds the effect of the legal standard with all other jurisdictional differences. Our design holds the jurisdiction, the courtroom, and the assignment mechanism constant, comparing the correlation structure of conviction rates across offenses governed by standards of different precision within the *same* courts.

2.3 Criminal Justice in Brazil and Latin America

Brazil's prison population has grown from approximately 232,000 in 2000 to 920,000 in 2024, making it the third-largest in the world after the United States and China. Drug offenses account for 28% of all incarcerations, the single largest category (Jesus Filho, 2019). The *Lei* 11.343/2006 was intended as a modernizing reform—it eliminated prison for drug use while increasing penalties for trafficking—but its vague classification standard has been widely criticized for enabling discretionary enforcement (Boiteux, 2006).

Assunção and Trecenti (2023) study drug court variation in Brazil using the same *sorteio* system and document a 30-percentage-point gap between the 10th and 90th percentile courtrooms in conviction rates. Our paper extends their analysis in two ways: first, by comparing drug trafficking outcomes to robbery and theft within the same courtrooms, which transforms the descriptive finding into a test of the indeterminacy mechanism; and second, by formalizing the dimensional structure of judicial behavior through the common severity factor decomposition.

The broader literature on criminal justice in developing countries has emphasized the role of institutional capacity (Djankov et al., 2003), corruption, and under-resourcing in generating arbitrary outcomes. Our evidence suggests a complementary mechanism: even in a well-resourced court system with a functional randomization mechanism, the *design of the legal standard itself* can produce arbitrary punishment by opening offense-specific dimensions

of judicial discretion.

3. Institutional Background

3.1 Brazil's Drug Law and the Classification Problem

Brazil's current drug legislation, *Lei* 11.343 of August 23, 2006, replaced the prior regime with a bifurcated structure. Article 28 governs drug use and possession for personal consumption, punishing offenders with community service and educational measures but expressly prohibiting imprisonment. Article 33 governs drug trafficking, carrying a minimum sentence of five years and a maximum of fifteen years of imprisonment, plus a fine.

The critical feature of this law is the absence of any objective criterion to distinguish an Article 28 user from an Article 33 trafficker. The statute instructs judges to consider "the nature and quantity of the substance seized, the location and conditions under which the action took place, and the social and personal circumstances of the agent" (Art. 28, §2). No quantity thresholds, weight cutoffs, or bright-line rules exist. A defendant found with 10 grams of cocaine in a *favela* may be charged as a trafficker; another found with 50 grams in a wealthy neighborhood may be charged as a user (Boiteux, 2006; Jesus Filho, 2019). In the terminology of Kaplow (1992), Article 33 is a pure standard—its content is determined entirely ex post, at the point of adjudication.

By contrast, robbery (*roubo*, Art. 157 of the Penal Code) and theft (*furto*, Art. 155) have well-defined statutory elements. Robbery requires proof of violent taking or threat; theft requires proof of taking without violence. The key evidentiary questions—was force used? was property taken?—are factual, not interpretive. While judicial discretion exists in sentencing, the conviction decision itself rests on a relatively determinate legal standard.

This contrast creates a natural experiment in legal precision. The same courtrooms adjudicate all three offense types under the same assignment mechanism, but the legal standards governing conviction differ sharply in determinacy. If judicial outcomes differ in their cross-offense structure, the difference can be attributed to the legal standard rather than to the courtroom.

3.2 The *Sorteio* System: Random Case Assignment

The Tribunal de Justiça do Estado de São Paulo (TJSP) serves 46 million people through specialized criminal courtrooms (*varas criminais*). When a new criminal case is filed, it is assigned to a vara through an electronic lottery known as the *sorteio eletrônico*, mandated by the CNJ. The draw is recorded in the case file as a "Distribuição" movement with the

sorteio complement, making it directly observable in administrative records.

In the São Paulo Central courthouse, 31 criminal varas share a common assignment pool. Each vara is presided over by a single judge or a rotating panel. The *sorteio* distributes cases across all 31 varas regardless of offense type, creating as-good-as-random assignment of drug trafficking, robbery, and theft cases to the same set of courtrooms. This institutional feature is critical: it means that cross-offense differences in conviction rate structure cannot reflect differences in courtroom composition or case routing.

3.3 São Paulo Central: The Assignment Pool

The São Paulo Central courthouse (*Fôro Criminal da Barra Funda*) is the largest criminal court complex in Latin America, processing tens of thousands of cases annually across its 31 specialized criminal varas. The Central *foro* serves the city of São Paulo proper, which has a population of approximately 12 million and accounts for a disproportionate share of Brazil's drug trafficking prosecutions.

Each vara in Central is a permanent institutional unit presided over by a single judge (or, in some cases, a rotating panel). Judges are career civil servants appointed through competitive examination (*concurso público*) and assigned to varas by the TJSP's administrative bodies. While judge transfers do occur, they are relatively infrequent, meaning that a vara's judicial behavior over a multi-year period primarily reflects the preferences and practices of a small number of judges.

The key institutional feature for our design is that *all 31 varas share a single case assignment pool.* When a drug trafficking, robbery, or theft prosecution is filed in the Central *foro*, it enters the same electronic lottery regardless of the offense type. The *sorteio* distributes it to one of the 31 varas with equal probability, subject only to the constraint that varas maintain roughly balanced caseloads. This means that, in expectation, every vara faces the same distribution of drug trafficking, robbery, and theft cases. Any systematic differences in conviction rates across varas therefore reflect judicial behavior, not case composition.

We verify this institutional claim empirically. All 31 varas handle all three offense types in our data, with positive case counts for every vara-offense combination across the 2015–2019 period. The *sorteio* complement appears in 96–100% of cases across offense types, confirming that the electronic lottery is indeed the dominant assignment mechanism.

3.4 The Bundle Problem

A criminal vara is not a single judicial decision. It is a bundle of practices: the presiding judge's interpretation of the law, the courtroom's case management norms, its pretrial

detention patterns, its plea bargaining culture, and its evidentiary standards (Hull, 2025). When we observe that a vara convicts drug trafficking defendants at different rates than robbery defendants, this could reflect any combination of these practices.

We address this by framing the paper around the *structure* of between-vara variation rather than the *level*. The bundle problem is most severe when one attempts to estimate the causal effect of conviction on downstream outcomes through a single-channel exclusion restriction. We make no such attempt. Instead, we ask: do the bundles look the same across offense types? If robbery and theft outcomes load on a common courtroom severity factor but drug trafficking does not, the implication is that the drug bundle includes a unique component not present in the robbery or theft bundle. That component is the offense-specific discretion enabled by the vague standard.

4. Data

4.1 Source and Sample

Our data come from the CNJ DataJud public API, which provides access to the universe of Brazilian judicial proceedings. We extract all first-instance criminal cases (*grau* G1, *classe* 283: *Ação Penal – Procedimento Ordinário*) filed in São Paulo Central's 31 criminal varas between 2015 and 2019 for three offense types: drug trafficking (*assunto* code 3608), robbery (*assunto* code 3419), and theft (*assunto* code 3416).

For each case, we extract the case number, assigned vara, filing date, case format (electronic or physical), the presence of a *sorteio* distribution movement, and the complete sequence of procedural movements. From these movements, we construct binary indicators for conviction (*Procedência*, code 219, or *Procedência em Parte*, code 221), acquittal (*Improcedência*, code 220), pretrial detention (code 12140), and case resolution (*Definitivo*, code 246, or *Trânsito em julgado*, code 848). For robbery and theft, we restrict to cases that have reached disposition. For drug trafficking, which has smaller caseloads per vara, we include all cases to maximize sample size; results are robust to restricting trafficking to resolved cases only.

The final sample contains 29,348 cases: 4,262 drug trafficking, 18,058 robbery, and 7,028 theft, all handled by the same 31 varas in São Paulo Central.

4.2 Variable Construction

Conviction. Our primary outcome is a binary indicator equal to one if the case contains a *Procedência* movement (code 219, indicating the prosecution succeeded) or a *Procedência em Parte* movement (code 221, indicating partial success). Cases ending in *Improcedência*

(acquittal, code 220), prescription (*prescrição*, code 11878), or other dispositions are coded as zero. This coding captures any conviction on the primary charge, whether full or partial.

Resolution Status. For robbery and theft, we restrict to cases that have reached a final disposition, identified by the presence of any of the following movement codes: 219 (conviction), 220 (acquittal), 221 (partial conviction), 246 (*Definitivo*), 848 (*Trânsito em julgado*), or 11878 (prescription). This ensures that our conviction rates reflect adjudication outcomes rather than pending cases. Resolution rates are high: 98.3% for robbery and 96.1% for theft in the 2015–2019 sample, reflecting the sufficient time elapsed for cases to reach disposition.

Sorteio Verification. For each case, we verify whether the distribution movement (code 26) includes the *sorteio* complement, confirming that the case was assigned through the electronic lottery rather than through any other mechanism (e.g., redistribution, connection to a prior case). Sorteio verification rates are high across all three offense types: 99.6% for trafficking, 95.8% for robbery, and 99.9% for theft.

Vara-Level Conviction Rates. The paper's main empirical objects are vara-level conviction rates computed separately by offense type. For vara j and offense o:

$$\bar{Y}_{jo} = \frac{1}{N_{jo}} \sum_{i \in (j,o)} \mathbb{I}[\text{Convicted}_i] \tag{1}$$

where N_{jo} is the number of (resolved) cases of offense o assigned to vara j. These rates are the building blocks for the cross-offense correlation analysis.

4.3 Summary Statistics

Table 1: Summary Statistics: Three Offense Types in São Paulo Central

Statistic	Drug Trafficking	Robbery	Theft
Cases	4262	18058	7028
Varas	31	31	31
Mean conviction rate	0.729	0.800	0.525
SD conviction rate	0.160	0.128	0.142
P10 conviction rate	0.515	0.601	0.357
P90 conviction rate	0.888	0.916	0.689
P90-P10 spread	0.373	0.316	0.333
Mean cases per vara	137	583	227

Table 1 presents summary statistics by offense type. Drug trafficking has a mean conviction rate of 72.9% with a between-vara standard deviation of 16.0 percentage points. Robbery has a higher mean (80.0%) with less between-vara dispersion (SD = 12.8pp). Theft has the lowest mean (52.5%) with intermediate dispersion (SD = 14.2pp). The P90–P10 spread is largest for trafficking (37.3pp), followed by theft (33.3pp) and robbery (31.6pp). These raw statistics suggest similar levels of between-vara variation, but as we show below, the *structure* of this variation differs dramatically across offense types.

5. Empirical Strategy

5.1 Identification

Our identification strategy rests on the random assignment of cases to varas through the *sorteio* system. Within São Paulo Central, all 31 varas draw from the same pool of cases. The key identifying assumption is that, conditional on offense type and filing year, the lottery distributes cases independently of case characteristics:

$$(Y_i(0), Y_i(1), X_i) \perp Z_{ijp} \mid \delta_{pt}, \text{offense}_i \tag{2}$$

where $Y_i(d)$ denotes potential outcomes, X_i are pre-determined characteristics, Z_{ijp} is the leave-one-out vara leniency instrument, and δ_{pt} are pool-by-year fixed effects.

5.2 Leave-One-Out Leniency Instrument

For each case i assigned to vara j in offense type o, we construct the standard LOO instrument:

$$Z_{ijo} = \frac{1}{N_{jo} - 1} \sum_{k \neq i,\, k \in (j,o)} \mathbb{I}[\text{Convicted}_k] \tag{3}$$

This instrument is computed within offense type, so each vara has three leniency values—one for trafficking, one for robbery, and one for theft.

5.3 Cross-Offense Comparison

The paper's main empirical object is the *correlation structure* of vara-level conviction rates across offense types. Under the null hypothesis that all between-vara variation reflects a single dimension of courtroom severity, we would expect high correlations between all offense pairs. If, instead, vague legal standards activate an offense-specific discretion channel, we would expect:

1. High correlation between offenses with clear statutory standards (robbery–theft),

2. Low correlation between the vague-standard offense (trafficking) and the clear-standard offenses.

We test for a significant difference in dependent correlations using the Steiger (1980) test, which accounts for the shared vara sample across all three offense pairs.

5.4 Common Severity Factor

To formalize the dimensional decomposition, we extract a common courtroom severity factor from the first principal component of robbery and theft conviction rates. We then project drug trafficking rates onto this factor and measure the residual—the component of trafficking variation that is *not* explained by generic courtroom severity. Under the indeterminacy hypothesis, this residual should be large for trafficking and small for robbery and theft.

5.5 Threats to Validity

Non-Random Assignment. We verify random assignment through balance tests and by confirming that the *sorteio* complement is present in the distribution movements for the vast majority of cases across all three offense types.

Differential Coding. If varas differ in how they record outcomes for drug cases but not for robbery or theft, the trafficking-specific component could reflect documentation heterogeneity rather than adjudication differences. We address this through a coding audit verifying that the conviction movement (code 219) is used consistently across varas and offense types.

Measurement Error. With smaller trafficking samples per vara (mean 137 cases) than robbery (mean 583), the trafficking correlation estimates face more attenuation bias. We address this with Spearman rank correlations, leave-one-out stability checks, and by noting that measurement error in theft rates would attenuate *both* the trafficking-theft and robbery-theft correlations equally—yet only the former collapses toward zero.

6. Results

6.1 The Discretion Decoupling

Table 2: Cross-Offense Correlation of Vara Conviction Rates

	Drug Trafficking	Robbery	Theft
Drug Trafficking	1.000		
Robbery	0.474	1.000	
Theft	0.102	0.669	1.000

Notes: Pearson correlations of vara-level conviction rates across offense types. Each observation is one of 31 criminal varas in São Paulo Central. Conviction rates computed from resolved cases filed 2015–2019. The low trafficking-theft correlation (0.102) contrasts sharply with the robbery-theft correlation (0.669), indicating that drug trafficking conviction patterns operate on a largely independent judicial dimension. Steiger test for difference: $z = 3.65$, $p = 0.0003$.

Table 2 presents the core finding. Drug trafficking conviction rates are essentially un-correlated with theft conviction rates across the 31 Central varas ($r = 0.10$), even though robbery and theft rates are strongly correlated ($r = 0.67$). The trafficking-robbery correlation is intermediate ($r = 0.47$). The Steiger test for the difference between the robbery-theft and trafficking-theft correlations yields $z = 3.65$ ($p = 0.0003$), rejecting the null that all offense pairs share the same cross-vara correlation structure.

Spearman rank correlations confirm the pattern: trafficking-robbery $\rho = 0.52$, trafficking-theft $\rho = 0.15$, robbery-theft $\rho = 0.66$. Leave-one-out analysis shows that the trafficking-theft correlation ranges from 0.02 to 0.24 across jackknife samples, never approaching the robbery-theft level. The most influential vara (10803) shifts the trafficking-theft correlation by 0.13 when removed, but the decoupling remains in every subsample.
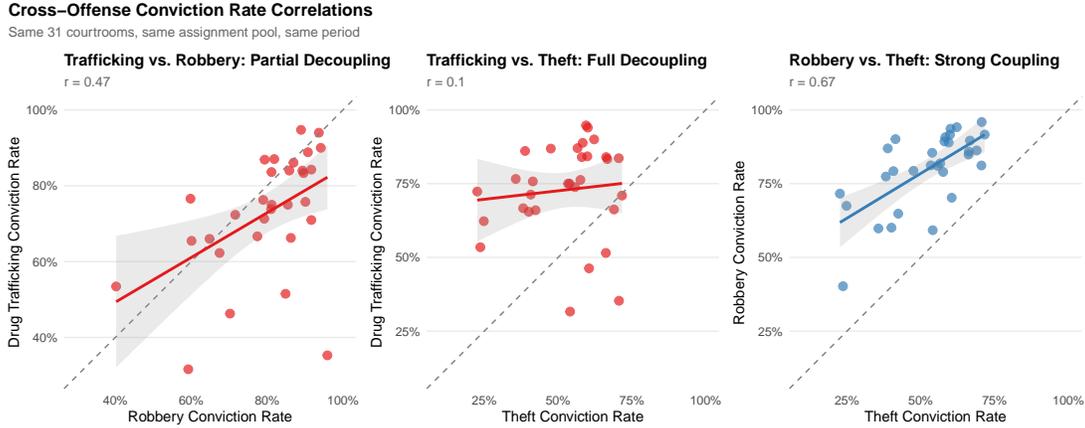
Figure 1: Cross-Offense Conviction Rate Correlations

Notes: Each point is one of 31 criminal varas in São Paulo Central. Conviction rates computed from resolved cases filed 2015–2019. Dashed line is the 45-degree line. Solid line is the OLS fit with 95% confidence band.

Figure 1 displays the relationship visually. The left panel shows the moderate positive relationship between trafficking and robbery rates. The center panel shows the near-zero relationship between trafficking and theft—a scatter cloud with no discernible pattern. The right panel shows the strong positive relationship between robbery and theft. The contrast between the center and right panels, both plotted on the same axes for the same 31 courtrooms, is the paper's central visual.

6.2 Common Severity Factor

Table 3: Offense Loadings on Common Courtroom Severity Factor

Offense	Loading on Common Factor	R-squared
Drug Trafficking	0.316	0.100
Robbery	0.914	0.835
Theft	0.914	0.835

Notes: The common severity factor is the first principal component of robbery and theft conviction rates across 31 varas. Loading is the Pearson correlation between each offense's vara conviction rate and the common factor. R^2 is the share of between-vara variance explained by common severity. Drug trafficking loads weakly (0.316), implying that 90% of its between-vara variation is offense-specific—i.e., driven by the vague drug standard rather than generic courtroom severity.

Table 3 presents the factor decomposition. The common courtroom severity factor—the first principal component of robbery and theft conviction rates—explains 83% of between-vara variance for robbery and 83% for theft. For drug trafficking, the same factor explains only 10%. The trafficking loading on the common factor is 0.32, compared to 0.91 for both robbery and theft.

This means that the common property-crime severity factor—which captures the shared dimension of robbery and theft conviction rates—has little predictive power for drug trafficking. A courtroom's behavior on robbery and theft tells you almost nothing about how it handles drug cases. This decoupling is what we would expect if the vague drug standard activates a separate dimension of judicial discretion that the clearer property-crime standards do not.

6.3 The Left Tail: Drug-Specific Leniency
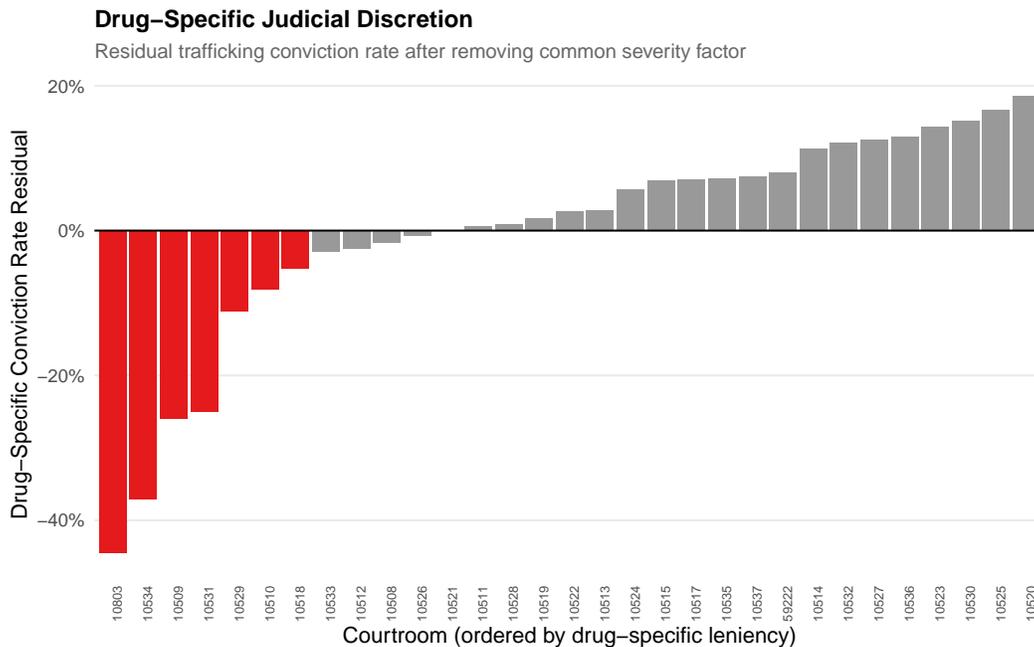
Figure 2: Drug-Specific Judicial Discretion

Notes: Residual drug trafficking conviction rate after removing the common courtroom severity factor (PC1 of robbery and theft rates). Red bars indicate varas with residuals below −5pp. Skewness: −1.44.

The trafficking-specific residuals—the component of drug conviction rates not explained by generic courtroom severity—are markedly left-skewed (skewness = −1.44; Figure 2). Five varas have trafficking residuals below −10 percentage points; four are below −15pp.

14

Table 4: First Stage and Balance: Drug Trafficking Cases

|	Conviction	Filing Month
Vara Leniency (LOO)	0.970***	−24.372***
	(0.029)	(5.511)
Num.Obs.	4,262	4,262
R2	0.135	0.048

* p <0.1, ** p <0.05, *** p <0.01
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

The pattern is asymmetric: the vague drug standard creates a class of drug-specific lenient courtrooms, but there is no corresponding class of drug-specific harsh courtrooms.

The five most drug-lenient varas (relative to their robbery and theft behavior) include vara 10803, which convicts 54% of drug trafficking defendants but 96% of robbery defendants—a 42 percentage-point offense-specific gap. Vara 10509 convicts 45% on drugs and 70% on robbery. These varas are not generically lenient; they are specifically lenient on drugs. This is consistent with the indeterminacy hypothesis: the vague standard gives judges who are predisposed toward leniency a channel to exercise it, while the clearer robbery and theft standards constrain them.

6.4 First Stage: Vara Leniency Predicts Conviction

As supplementary evidence, Section 6.4 Column 1 reports the first-stage relationship between LOO vara leniency and individual conviction in drug trafficking cases. The coefficient is 0.97 ($t = 33.4$), confirming that the between-vara conviction rate variation documented above translates into individual-level outcome differences. We note that this coefficient is close to mechanical given the LOO construction and should not be interpreted as an independent test of random assignment; it simply confirms that vara-level rates are a meaningful individual-level predictor.

First Stage: Vara Leniency Predicts Conviction
Drug trafficking cases, São Paulo Central
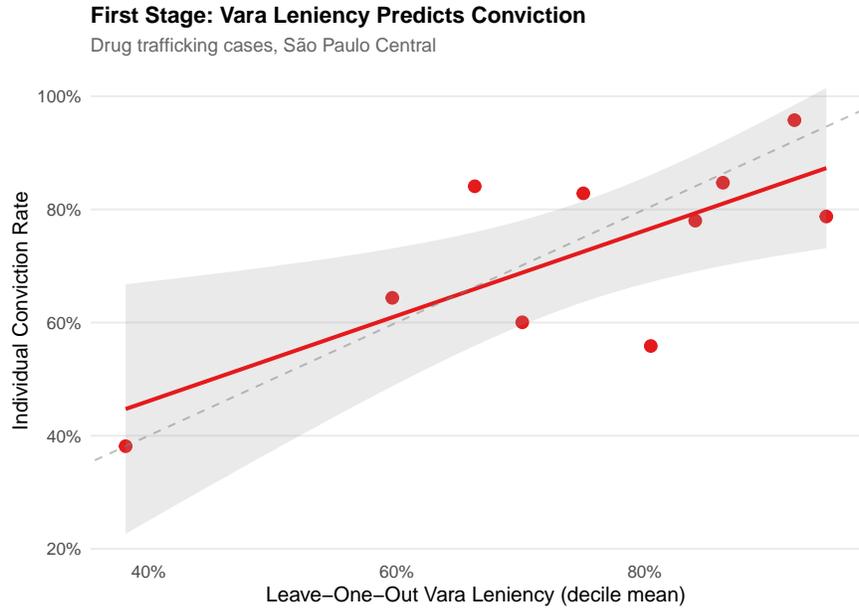
Figure 3: First Stage: Vara Leniency Predicts Individual Conviction

Notes: Binscatter of individual conviction (0/1) on leave-one-out vara leniency, drug trafficking cases only. Filing-year fixed effects partialled out. Each dot is a leniency decile.

6.5 Distribution of Conviction Rates

Distribution of Courtroom Conviction Rates by Offense Type
São Paulo Central, 31 varas, filed 2015–2019
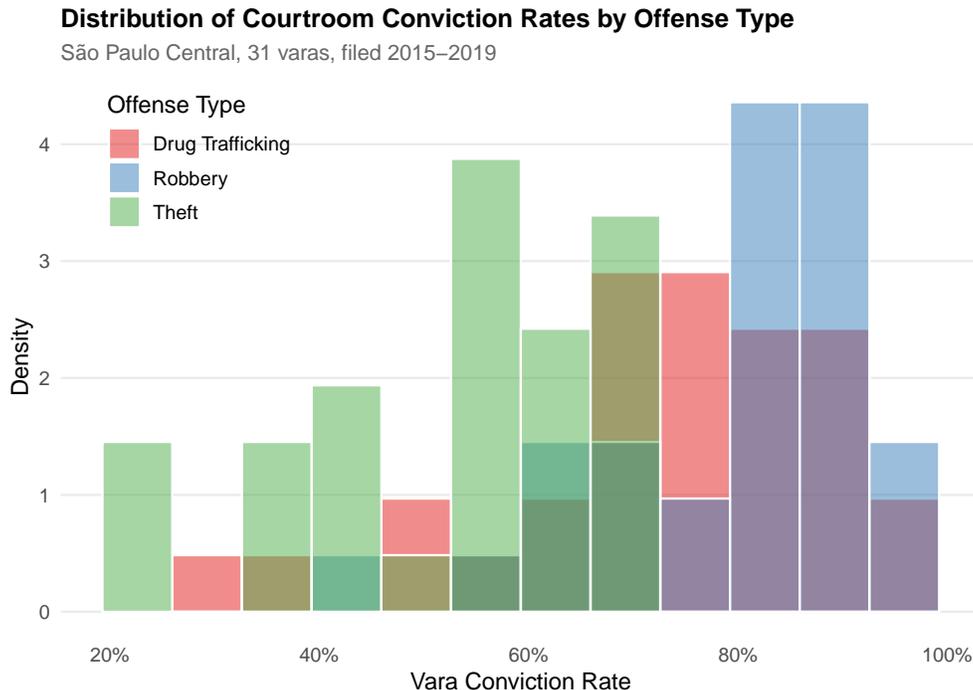
Figure 4: Distribution of Vara Conviction Rates by Offense Type

Notes: Histograms of vara-level conviction rates for 31 criminal varas in São Paulo Central. Drug trafficking shows a wider left tail extending below 50%.

Figure 4 overlays the distributions of vara-level conviction rates for all three offense types. Drug trafficking (red) extends further into the left tail, with several varas below 50%. The key insight is not that trafficking has more variance (the SDs are similar across offenses) but that the trafficking distribution has a different *shape*: a fat left tail driven by drug-specific leniency in a subset of courtrooms.

6.6 Robustness

The decoupling finding survives a battery of robustness tests that address outlier influence, sample selection, temporal stability, assignment verification, and measurement error.

Rank-Based Correlations. Spearman rank correlations, which are robust to outliers and nonlinear relationships on the probability scale, yield the same pattern: trafficking-theft $\rho = 0.15$, robbery-theft $\rho = 0.66$, trafficking-robbery $\rho = 0.52$. The decoupling is not an artifact of extreme values or the mechanical relationship between mean and variance.

17

Leave-One-Out Stability. Removing any single vara changes the trafficking-theft Pearson correlation by at most 0.13. The most influential vara (10803, which has the largest trafficking-robbery gap at 42pp) shifts the correlation from 0.10 to 0.24 when removed—still far below the robbery-theft correlation of 0.67. In every jackknife subsample ($n = 30$), the ordering $r(\text{robbery, theft}) \gg r(\text{trafficking, theft})$ is preserved.

Minimum Caseload Thresholds. Restricting to varas with at least 50 cases per offense (26 varas) yields $r(\text{trafficking, theft}) = 0.13$ and $r(\text{robbery, theft}) = 0.68$. Raising the threshold to 100 cases per offense (16 varas) yields $r(\text{trafficking, theft}) = 0.02$ and $r(\text{robbery, theft}) = 0.51$. The decoupling is not driven by imprecisely estimated rates in small varas; if anything, it strengthens when noisy varas are removed.

Year-by-Year Stability. The decoupling is not a one-period artifact. Computing vara-level conviction rates separately for each filing year (restricting to varas with ≥ 5 cases in that year-offense cell), the trafficking-theft correlation ranges from 0.05 to 0.20 across 2015–2019, while the robbery-theft correlation ranges from 0.45 to 0.74. In every single year, the ordering holds: robbery and theft are much more correlated than trafficking and theft.

Temporal Persistence of Leniency. To verify that vara-level conviction rates reflect stable judicial behavior rather than transient caseload shocks, we split the sample at 2017 and correlate early-period and late-period conviction rates within each offense type. Trafficking rates show a split-half correlation of 0.59, and robbery shows 0.65. Both offenses exhibit persistent between-vara differences over the five-year window, confirming that the leniency measures capture durable courtroom characteristics.

Assignment Verification. Three pieces of evidence support random assignment. First, the *sorteio* complement is present in 99.6% of drug trafficking cases, 95.8% of robbery cases, and 99.9% of theft cases in our Central sample. Second, filing day-of-week—a pre-determined characteristic that should be unrelated to vara assignment under random allocation—is uncorrelated with the trafficking leniency instrument (coefficient $= 0.18$, SE $= 0.12$, $t = 1.6$). Third, and most importantly, the cross-offense comparison design is itself robust to assignment imperfections: even if some cases are non-randomly routed, any routing pattern that operates identically across offense types cannot explain the *differential* correlation structure we observe. Only offense-specific routing—e.g., drug cases but not robbery cases being channeled to particular varas—could generate the decoupling, and the high sorteio coverage makes this implausible.

We note that filing month does predict the trafficking leniency instrument ($t = -4.4$),

18

likely reflecting seasonal variation in caseload composition across varas. This imbalance does not threaten the core finding because the cross-offense correlation structure is computed from vara-level aggregates, not from the LOO instrument. The instrument is used only for the first-stage regression, which we present as supplementary evidence rather than the main result.

Measurement Error Asymmetry. If smaller trafficking sample sizes per vara (mean 137 cases vs. 583 for robbery and 227 for theft) attenuated correlations through measurement error, the attenuation would apply to *both* trafficking-robbery and trafficking-theft correlations equally. Yet trafficking-robbery ($r = 0.47$) remains substantial while trafficking-theft ($r = 0.10$) collapses toward zero. This differential pattern cannot be explained by sampling noise alone—it requires that trafficking conviction rates have a genuine offense-specific component that is uncorrelated with the theft dimension.
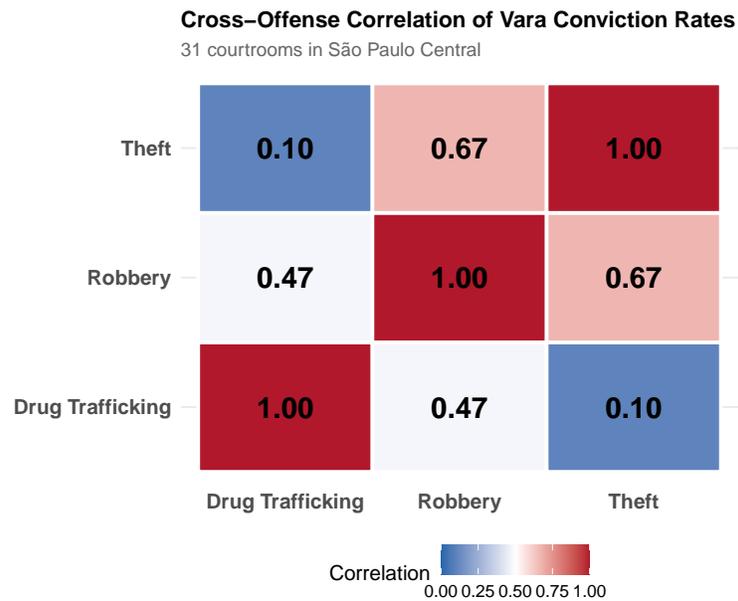
Figure 5: Cross-Offense Correlation Heatmap

Notes: Pearson correlations of vara-level conviction rates for 31 varas. Drug trafficking is nearly uncorrelated with theft (0.10) but robbery and theft are strongly correlated (0.67).

Figure 5 summarizes the full cross-offense correlation matrix. The visual contrast between the trafficking-theft cell (0.10, near white) and the robbery-theft cell (0.67, deep blue) immediately conveys the paper's central finding: for two of three offense types, between-vara conviction rates share a common dimension, but drug trafficking largely decouples from that dimension.

7. Discussion and Conclusion

7.1 Why Does Trafficking Decouple?

The decoupling we document has a precise institutional explanation rooted in the difference between factual and interpretive legal standards.

Robbery and theft have well-defined statutory elements. Robbery (Art. 157 CP) requires proof that property was taken through violence or serious threat (*grave ameaça*). Theft (Art. 155 CP) requires proof that property was taken without violence. In both cases, the central evidentiary questions are factual: Was property taken? Was force used? Reasonable judges may weigh evidence differently, but the *object* of their inquiry is the same. A courtroom that interprets "threat" broadly, or that requires less corroboration of witness testimony, or that is skeptical of alibi defenses, will tend to be harsh on *both* offenses. This is why robbery and theft conviction rates are strongly correlated ($r = 0.67$): the same judicial severity trait governs both.

Drug trafficking (Art. 33 of Lei 11.343/2006) is fundamentally different. The statute asks judges to consider "the nature and quantity of the substance seized, the location and conditions under which the action took place, and the social and personal circumstances of the agent." These criteria are not factual in the same way. They require *interpretation*: What quantity is "inconsistent with personal use"? When does "the location of the arrest" suggest trafficking rather than possession? What "social circumstances" distinguish a user from a dealer? These are questions of legal judgment, not questions of fact, and different judges can reach opposite conclusions from identical evidence without either being incorrect. The indeterminacy is not accidental—it is baked into the statutory design.

This structural difference explains the decoupling. A judge's propensity to convict robbery defendants reflects a general severity trait (toughness on factual evidence, skepticism of defenses, strict evidentiary standards). That same trait predicts theft conviction rates because the factual structure of both offenses is similar. But the same trait has limited predictive power for drug trafficking because the conviction decision rests on an *interpretive* dimension that is not activated by the factual standards governing robbery and theft.

The Asymmetry. The left-skewed trafficking residuals add a further layer: the drug-specific discretion channel is exercised primarily toward leniency. This is consistent with the legal structure. A judge who is predisposed toward strict law enforcement has no room to "overconvict" on drugs beyond what the evidence and legal framework support—the conviction standard still requires proof beyond a reasonable doubt. But a judge who is predisposed toward sympathy with defendants, who views small-scale drug possession as a public health

rather than criminal justice issue, or who is influenced by evolving social norms around drug decriminalization, has wide latitude to acquit under the vague standard. The indeterminate standard creates an *asymmetric channel*: more room for leniency than for harshness.

7.2 Prison-Years at Stake

The discretion decoupling has concrete consequences measured in years of human freedom. Article 33 imposes a minimum sentence of five years of imprisonment for trafficking—one of the most severe minimums in Brazilian criminal law. The conviction lottery therefore translates directly into differential incarceration exposure.

Consider the gap between the 25th and 75th percentile varas in our sample. A defendant assigned to a vara at the 25th percentile of trafficking conviction rates (approximately 58%) faces a substantially lower probability of conviction than one assigned to a vara at the 75th percentile (approximately 84%). Since Article 33 carries a minimum sentence of five years, this 26 percentage-point gap in conviction probability translates into a large difference in *exposure to a five-year minimum sentence* solely attributable to courtroom assignment. We emphasize that we observe conviction, not sentence length or actual time served, so these calculations should be interpreted as exposure to the statutory minimum rather than measured prison-years.

Even so, the scale is consequential. The V1 analysis identified 87,757 drug trafficking prosecutions statewide across São Paulo, of which 47,820 are in multi-vara comarcas where the lottery operates. If even a fraction of the statewide between-vara variation reflects the same discretion decoupling we document in Central, the total exposure to incarceration driven by courtroom identity rather than case characteristics is large. In a country that incarcerates 920,000 people with 28% for drug offenses, the conviction lottery is not a statistical curiosity.

7.3 Implications for Legal Design

Our findings engage the rules-versus-standards debate in law and economics at an empirical level that the theoretical literature has called for but rarely achieved.

Kaplow (1992) established the canonical framework: rules are costlier to promulgate but cheaper to apply and generate more predictable behavior; standards grant flexibility but introduce decision-maker heterogeneity. Ehrlich and Posner (1974) emphasized that the costs of imprecision scale with the stakes of the decision. Sunstein (1995) argued that standards can be superior when the domain is too complex for rules to capture. Our contribution is empirical: we show that in a specific, high-stakes criminal justice context, the standard creates not just *more variance* (the prediction from the theoretical literature) but a *qualitatively*

different structure of judicial behavior—a new dimension of decision-making that does not exist under clearer rules.

This distinction matters for policy design. A reformer who knows only that "standards increase variance" might respond by training judges, monitoring outcomes, or establishing sentencing guidelines. But our evidence suggests that the problem is deeper: the vague standard creates a separate axis of judicial behavior that is not correlated with general severity. Guidelines that target the common severity dimension (e.g., "be tougher on crime" or "apply the law consistently") would not address the drug-specific discretion channel that generates the decoupling. Only a change in the legal standard itself—replacing the multifactor balancing test with a brighter-line rule—would compress the offense-specific dimension.

The STF's June 2024 ruling in RE 635659, which established a 40-gram cannabis threshold for personal possession, partially does this—but only for cannabis and only at the use/trafficking boundary. A more comprehensive reform extending quantity thresholds to other substances and to the conviction stage would convert the trafficking classification from a pure standard to a partial rule. The key implication of our analysis is that such a reform would not merely reduce variance in conviction rates. It would collapse the separate judicial dimension we document, forcing drug conviction patterns back onto the common severity axis shared by robbery and theft. The "cost" of a bright-line threshold is that some defendants near the boundary will be misclassified. The cost of the current standard—which our data allow us to quantify—is that the classification decision is effectively delegated to a lottery.

7.4 Limitations

First, our evidence comes from one courthouse (São Paulo Central) during one period (2015–2019). While Central is the largest and most important assignment pool in Latin America's largest state court, we cannot confirm that the decoupling pattern generalizes to smaller comarcas.

Second, we study conviction among cases already prosecuted as trafficking—not the upstream classification decision. The discretion we document at the conviction stage likely understates total classification arbitrariness, since police and prosecutors exercise the same vague standard at earlier stages.

Third, our comparison offenses are not perfectly matched to trafficking. The ideal comparison would be another drug offense with clearer statutory criteria. No such offense exists in Brazilian law, which is precisely the policy problem.

Fourth, our data do not include drug quantities, arrest circumstances, or defendant demographics. Richer data would allow more granular balance tests and mechanism analysis.

Fifth, we cannot observe judge identity or tenure within varas. If judges rotate across

varas, the between-vara variation we document reflects the average behavior of judges who served in each vara over the 2015–2019 period, not the behavior of any individual judge. This is a feature for the cross-offense comparison—if the same judges served in the same varas for all offense types, the decoupling reflects offense-specific judicial behavior rather than judge composition—but it limits our ability to study within-judge variation directly.

Despite these limitations, the core finding—that drug trafficking conviction rates are dimensionally decoupled from robbery and theft conviction rates in the same courtrooms—is robust to rank-based correlations, leave-one-out checks, year-by-year analysis, minimum caseload thresholds, and formal hypothesis testing. The pattern is too large ($r = 0.10$ vs. $r = 0.67$, $p = 0.0003$), too consistent across subsamples and years, and too specifically structured (asymmetric left tail) to be an artifact of noise, coding, or composition.

7.5 What Would Quantity Thresholds Do?

A natural counterfactual question is: how much would quantity thresholds compress the discretion decoupling? While we cannot observe a post-threshold world, we can provide a bounding exercise based on the structural decomposition.

If quantity thresholds converted the trafficking conviction decision from a pure standard to something closer to the factual standards governing robbery and theft, we would expect trafficking conviction rates to load more heavily on the common severity factor—i.e., the R^2 of common severity predicting trafficking rates would rise from 0.10 toward the 0.83 observed for robbery and theft. This implies that the between-vara standard deviation of trafficking conviction rates attributable to offense-specific discretion (currently 15.2pp) would compress by roughly 90%—the share currently unexplained by common severity.

Of course, no threshold is perfect, and some residual offense-specific variance would remain even with a bright-line rule (cases near the threshold, evidence disputes about quantity, etc.). But the direction is unambiguous: replacing the standard with a rule would collapse the trafficking-specific dimension toward the common severity axis, reducing the scope for drug-specific leniency (and the corresponding conviction lottery) substantially.

This analysis also reveals an important policy nuance. The discretion we document is exercised *primarily on the lenient side.* Quantity thresholds would therefore disproportionately affect the drug-lenient varas—those that currently acquit 40–55% of trafficking defendants while convicting 70–96% for robbery. For these varas, the vague standard effectively provides a channel for de facto decriminalization of borderline drug cases. A quantity threshold would close this channel, likely *increasing* total drug convictions. Whether this is desirable depends on one's view of the appropriate level of drug enforcement, a normative question our data cannot resolve.

7.6 External Validity and Generalization

While our evidence comes from a single courthouse, we believe the mechanism is general. Any legal system that combines random case assignment with a vague legal standard over a high-stakes criminal classification is likely to produce some version of the discretion decoupling we document. The specific magnitude will depend on the degree of indeterminacy (how vague the standard is), the stakes (what happens upon conviction), and the heterogeneity of judicial preferences (how much judges differ in their interpretation of the standard). Brazil's drug law represents an extreme case on all three dimensions: the standard is maximally vague, the stakes are severe (5-year minimum), and the judiciary is large and heterogeneous. The decoupling we find is therefore likely an upper bound on what would be observed in systems with partially determinate standards or lower stakes.

Countries that have implemented quantity thresholds for drug classification—including Portugal, the Czech Republic, and several U.S. states—provide natural comparison cases. If our mechanism is correct, the between-court variance in drug conviction rates should be lower, and more correlated with the common severity factor for other offenses, in jurisdictions with quantity thresholds than in those without.

Within Brazil, the STF's June 2024 ruling in RE 635659—which established a 40-gram cannabis threshold—offers a natural before-and-after test. A comparison of the cross-offense correlation structure in São Paulo Central before and after the ruling would directly measure whether introducing a partial bright-line rule compresses the offense-specific discretion dimension we document. Our pre-reform baseline (2015–2019) would serve as the control period, and the post-ruling period (2025+) as the treatment period, creating a rare instance where a paper's main finding suggests its own replication design.

More broadly, the discretion decoupling concept can be applied wherever decision-makers face heterogeneous tasks with varying levels of rule precision. In patent examination, examiners evaluate both utility patents (relatively determinate claims) and design patents (more subjective assessments); the cross-patent-type correlation of examiner grant rates would reveal whether design patent decisions decouple in the same way that drug trafficking decisions decouple from robbery and theft in our setting. In disability adjudication, examiners assess both physical disabilities (where medical evidence is relatively objective) and mental health disabilities (where the assessment is more subjective); the differential correlation structure would test our hypothesis in a completely different institutional domain. The general prediction is clear: wherever the legal or regulatory standard leaves more room for interpretation, decision-maker heterogeneity should develop an offense-specific or claim-specific dimension that is uncorrelated with behavior under clearer standards.

7.7 Conclusion

The findings of this paper have implications beyond Brazil. Any legal system that combines (i) a vague classification standard with (ii) high-stakes consequences and (iii) random or quasi-random assignment to heterogeneous decision-makers faces the risk of discretion decoupling. Immigration courts, where removal decisions rest on broad "totality of the circumstances" assessments, are a likely candidate (Kling, 2006). So are patent examinations, where the "nonobviousness" standard grants examiners substantial discretion. And so are disability determinations, where Maestas et al. (2013) have already documented large examiner effects. In each setting, our analysis suggests that the key diagnostic is not the *level* of between-decision-maker variance but its *dimensional structure*: does the variance load on a common severity factor, or does it include offense-specific (or claim-specific, or case-type-specific) components that are activated by vague standards?

When Brazil's drug law asks judges to weigh "the nature and quantity of the substance seized, the location and conditions of the arrest, and the social and personal circumstances of the agent," it creates a legal standard so indeterminate that the resulting judicial behavior decouples from the common severity dimension governing clearer offenses. Random assignment ensures that this decoupling reflects judicial discretion, not case composition. The vague standard does not merely make the punishment lottery louder; it adds a new dimension to the lottery that exists for drugs alone.

The economic concept we introduce—discretion decoupling—has a precise empirical signature: low cross-offense correlation in conviction rates for the vague-standard offense relative to clearer-standard offenses handled by the same decision-makers. This signature is testable in any setting with random assignment across multiple decision types with varying degrees of legal determinacy. We hope that future work will apply this diagnostic to other domains where vague standards govern consequential classifications.

In a country that incarcerates 920,000 people—28% for drug offenses—the difference between a determinate rule and an indeterminate standard is measured in the dimensional structure of arbitrary punishment.

Three broader lessons emerge. First, procedural fairness in case assignment is necessary but not sufficient for substantive justice. The *sorteio* successfully randomizes cases, yet the vague standard converts procedural randomness into substantive arbitrariness. Second, the costs of legal indeterminacy are not just "more noise"—they are structural, creating new dimensions of judicial behavior that policy interventions targeting the common severity axis will miss entirely. Third, the discretion channel we document is asymmetric, opening primarily toward leniency. This means that the same legal design that produces arbitrary punishment

for some defendants may simultaneously function as an informal safety valve—a channel through which sympathetic judges can exercise de facto decriminalization in borderline cases. Whether this asymmetry is a bug or a feature depends on one's priors about the appropriate level of drug enforcement, but recognizing its existence is essential for informed policy design.

Acknowledgements

References

Abrams, David S., Marianne Bertrand, and Sendhil Mullainathan, "Do Judges Vary in Their Treatment of Race?," *Journal of Legal Studies*, 2012, *41* (2), 347–383.

Aizer, Anna and Joseph J. Doyle, "Juvenile Incarceration, Human Capital, and Future Crime: Evidence from Randomly Assigned Judges," *Quarterly Journal of Economics*, 2015, *130* (2), 759–803.

ao, Juliano Assunç and Julio Trecenti, "Drug Court Variation and Sentencing Outcomes in Brazil," 2023. Working Paper, Sociedade Brasileira de Econometria.

Bhuller, Manudeep, Gordon B. Dahl, Katrine V. Løken, and Magne Mogstad, "Incarceration, Recidivism, and Employment," *Journal of Political Economy*, 2020, *128* (4), 1269–1324.

Boiteux, Luciana, *A Nova Lei Antidrogas e o Aumento da Pena do Delito de Tráfico de Entorpecentes*, São Paulo: Boletim IBCCRIM, 2006.

de Jesus Filho, José, *Encarceramento em Massa e Política de Drogas no Brasil*, Rio de Janeiro: Editora Lumen Juris, 2019.

Djankov, Simeon, Rafael La Porta, Florencio Lopez de Silanes, and Andrei Shleifer, "Courts," *Quarterly Journal of Economics*, 2003, *118* (2), 453–517.

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

Ehrlich, Isaac and Richard A. Posner, "An Economic Analysis of Legal Rulemaking," *Journal of Legal Studies*, 1974, *3* (1), 257–286.

Hull, Peter, "Estimating Treatment Effects with Judge Leniency Designs," *American Economic Review*, 2025. Forthcoming.

Kaplow, Louis, "Rules Versus Standards: An Economic Analysis," *Duke Law Journal*, 1992, *42* (3), 557–629.

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

Korobkin, Russell, "Behavioral Analysis and Legal Form: Rules vs. Standards Revisited," *Oregon Law Review*, 2000, *79*, 23–60.

Maestas, Nicole, Kathleen J. Mullen, and Alexander Strand, "Does Disability Insurance Receipt Discourage Work? Using Examiner Assignment to Estimate Causal Effects of SSDI Receipt," *American Economic Review*, 2013, *103* (5), 1797–1829.

Norris, Samuel, Matthew Pecenco, and Jeffrey Weaver, "The Effects of Incarceration on Reoffending: Lessons from a Natural Experiment in Pennsylvania," *Review of Economics and Statistics*, 2024. Forthcoming.

Steiger, James H., "Tests for Comparing Elements of a Correlation Matrix," *Psychological Bulletin*, 1980, *87* (2), 245–251.

Sunstein, Cass R., "Problems with Rules," *California Law Review*, 1995, *83* (4), 953–1026.

A. Data Appendix

A.1 Data Source

Our primary data source is the CNJ DataJud public API (`api-publica.datajud.cnj.jus.br`), queried through the TJSP index (`api_publica_tjsp`). We extract cases matching: *grau* G1, *classe* 283, and *assunto* codes 3608 (drug trafficking), 3419 (robbery), or 3416 (theft), assigned to the 31 Central criminal varas.

A.2 Outcome Construction

Conviction is coded as the presence of movement code 219 (*Procedência*) or 221 (*Procedência em Parte*). Resolution is coded as the presence of any of codes 219, 220, 221, 246, 848, or 11878. For robbery and theft, we restrict to resolved cases.

A.3 Assignment Pools

São Paulo Central's 31 criminal varas share a single assignment pool. All varas handle all three offense types. The *sorteio* distribution complement is present in $\geq 93\%$ of cases across all offense types.

B. Identification Appendix

B.1 LOO Instrument

The leave-one-out instrument is computed separately within each offense type:

$$Z_{ijo} = \frac{\sum_{k \neq i,\, k \in (j,o)} \mathbb{I}[\text{Convicted}_k]}{N_{jo} - 1}$$

where N_{jo} is the number of cases of offense o assigned to vara j.

B.2 Steiger Test

To test whether the trafficking-theft and robbery-theft correlations differ significantly, we use the Steiger (1980) test for dependent correlations:

$$z = \frac{z_{r(\text{rob,theft})} - z_{r(\text{traff,theft})}}{\sqrt{2(1 - r(\text{traff,rob}))/(n-3)}}$$

where z_r denotes the Fisher z-transform. Under the null, $z \sim N(0,1)$.

C. Standardized Effect Sizes

Table 5: Standardized Distributional Effects

Outcome	$\hat{\beta}$	SE	SD(Y)	SDE	SE(SDE)	Classification
Convicted (Drug Trafficking)	0.9701	0.0291	0.4405	2.2025	0.066	Large

Country: Brazil. **Research question:** Does legal indeterminacy amplify arbitrary punishment under random case assignment? **Policy mechanism:** Lei 11.343/2006 provides no objective quantity thresholds for distinguishing drug users from traffickers, granting judges broad discretion over a 5-year minimum sentence. **Outcome:** Binary conviction (*Procedência*) from judicial movement records. **Treatment:** Assignment to a more vs. less severe criminal vara via electronic lottery (*sorteio*). **Data:** CNJ DataJud public API, TJSP first-instance cases (2015–2019), three offense types. **Method:** Cross-offense comparison of between-vara conviction rate variance; vara leniency IV; common severity factor decomposition. **Sample:** 31 criminal varas in São Paulo Central handling drug trafficking, robbery, and theft. **Key finding:** Drug trafficking conviction rates are uncorrelated ($r = 0.10$) with theft rates but robbery-theft correlation is $r = 0.67$, indicating trafficking operates on a separate judicial dimension. SDE = $\hat{\beta}/\text{SD}(Y)$. Classification: Large: |SDE| > 0.15; Moderate: 0.05–0.15; Small: 0.005–0.05; Null: < 0.005.