Fight like a Woman: Domestic Violence and Female Judges in Brazil*

Helena Laneuville† Vitor Possebom‡

First Draft: March 2024; This Draft: July 2024
Please click here for the most recent version

Abstract

We investigate the impact of judges' gender on the outcome of domestic violence cases. Using data from São Paulo, Brazil, between 2010 and 2019, we compare conviction rates by judge's gender and find that a domestic violence case assigned to a female judge is 31% (10 p.p.) more likely to result in conviction than a case assigned to a male judge with similar career characteristics. To show that this decision gap rises due to different gender perspectives about domestic violence and not because female judges are stricter than their male counterparts in all rulings, we compare it against the gender conviction-rate gap in similar types of crime. We find that the gender conviction-rate gap for domestic violence cases is significantly larger than the same gap for other misdemeanor cases (3 p.p. larger) and for other physical assault cases (8 p.p. larger). Furthermore, we find evidence that at least two channels explain this gender conviction-rate gap for domestic violence cases: gender-based differences in evidence interpretation and gender-based sentencing criteria. Lastly, we find that this gender conviction rate has no significant impact on the probability of appeals, ruling reversals or recidivism.

Keywords: Domestic Violence, Judicial Decisions, Gender Bias

JEL Codes: D73, J16, K42

*We would like to thank Fernanda Estevan, John Eric Humpries, Cormac O'Dea, Laura Schiavon, Henrik Sigstad, and Winnie van Dijk for helpful suggestions.

†Instituto Mobilidade e Desenvolvimento Social - Imds. Email: helena.garcia@imdsbrasil.org

‡Sao Paulo School of Economics - FGV. Email: vitor.possebom@fgv.br

1 Introduction

Violence against women is frequently perpetrated inside the home by the ones closest to them. The United Nations Office on Drugs and Crime (2023) estimates that 55% of all homicides against women worldwide in 2022 were committed by intimate partners or other family members. As this violence is a criminal offense, key components of the fight against domestic violence are the Criminal Court System and how court judges approach this type of crime (Beecher-Monas, 2001). For instance, the Human Rights Watch (1991) argues that, in the 1980s and early 1990s, Brazilian judges failed to analyze crimes of domestic violence against women in a non-discriminatory manner.[1] Moreover, Kafka et al. (2019) finds that American judges in North Carolina use gender stereotypes to argue that some women want to "game the system" with "frivolous cases" when requesting protective orders against their abusive partners.

Since discriminatory gender stereotypes may vary depending on the decision maker's individual characteristics, we investigate the impact of judges' subjectivity on the outcome of domestic violence cases. In particular, we focus on the effect of a judge's gender on their decisions, documenting a sizeable and significant gender conviction-rate gap specific to domestic violence cases. We also analyze two possible forces driving this difference. At the end, we discuss the consequences of this gender conviction-rate gap beyond the Trial Court, analyzing the outcome of domestic violence cases in the Appeals Court and the recidivism behavior of these cases' defendants. We find that the Trial Judge's gender has no effect on these outcomes.

To do this, we collect data from São Paulo, Brazil, between 2010 and 2019 and compare conviction rates by the presiding judge's gender after controlling for the judge's career characteristics. Case characteristics are independent of the presiding judge's gender because cases are randomly allocated to judges within each court district in Brazil. Hence, the gender

[1] In the years since, Brazilian authorities have made important legal and institutional changes to address domestic violence crimes. The most prominent of these is the Maria da Penha Law, enacted in 2006, which characterizes domestic violence as a specific type of criminal offense. Ferraz and Schiavon (2022) find that this law reduced female homicides due to household aggression by 9 percent, with stronger effects for less educated and black women.

conviction-rate gap captures different decision criteria by judges of different genders.

In this analysis, we find that a domestic violence case assigned to a female judge is 31% (10 p.p) more likely to result in conviction than a case assigned to a male judge with similar career characteristics. Since this decision gap may be due to differences between male and female judges regardless of their perspectives on domestic violence, we compare this gap to the gender decision gaps for two other types of crime: misdemeanors and physical assault cases. Because we focus on domestic violence crimes that do not result in death or long-term physical impairment, misdemeanors are comparable to domestic violence cases in their sentence severity (according to Brazilian Law). Physical assault cases are comparable to domestic violence cases because both involve some sort of physical aggression.

We find that the gender conviction-rate gap for domestic violence cases is more than twice the size of that of other misdemeanors. Furthermore, while there is a gender conviction-rate gap for domestic violence cases, there are no differences between female and male judges when analyzing other types of physical assault. Consequently, the gender conviction-rate gap for domestic violence cases likely arises due to different perspectives about this type of crime rather than because female judges are, in general, more punitive than their male counterparts.

For this reason, we classify this type of gap as an in-group bias according to the definition given by Shayo and Zussman (2011). We do so because victims of domestic violence are overwhelmingly female, and members of this group (female judges) act differently from members of the other group (male judges) when analyzing this type of crime specifically.

We then investigate two possible drivers of this in-group bias. They were proposed by Boyd et al. (2010) and are known as the representational account and the informational account. The first one argues that female judges act as representatives of their group and tend to protect it more intensely than male judges. The second one argues that female judges process the information contained in domestic violence cases in a unique way. Our evidence suggests that both accounts drive the in-group bias in domestic violence cases in São Paulo, Brazil.

To investigate if the salience of gender identity matters (representational account), we test if the gender difference in strictness is more pronounced in intimate partner violence cases. To do so, we explore each case's full sentence and estimate whether female judges are more likely to convict a defendant for a domestic violence offense and simultaneously flag the case as associated with intimate partner violence. We find that the gender conviction-rate gap is 37.5% higher (9 p.p.) in cases in which a female judge may identify more strongly with the victim than a male judge and statistically null in cases that are not flagged as intimate partner violence offenses. This difference is associated with the representational account of the in-group bias because the incidence and tolerance of this type of offense are strongly associated with patriarchal gender norms involving partners (Heise and Kotsadam, 2015; Tur-Prats, 2019; González and Rodríguez-Planas, 2020).

To test for gender differences in the interpretation of the evidence provided in domestic violence cases (informational account), we compare the gender conviction rate between cases in which the defendant was caught by the police while committing the crime and cases without this objective piece of evidence. We find that male and female judges make similar decisions when analyzing cases where defendants are caught in the act, while they differ when making decisions about cases that rely on more subjective evidence. Hence, female judges process the information contained in domestic violence cases differently from their male peers, as suggested by the informational account of the in-group bias.

Lastly, we discuss the consequences of the gender conviction-rate gap beyond the Trial Court. To do so, we analyze two types of consequences of a Trial Judge's sentences. First, we look at the outcome of domestic violence cases in the Appeals Court, analyzing whether the Trial Judge's gender impacts how frequently there is an appeal and how frequently their sentence is reversed. Second, we look at the future behavior of defendants in domestic violence cases, analyzing whether the Trial Judge's gender impacts the defendants' criminal recidivism after the final sentence.

The impact of stricter female judges on appeals and ruling reversals in domestic violence cases is theoretically ambiguous. On the one hand, the positive gender conviction-rate gap may lead to female judges facing a larger reversal probability because sentences by stricter judges are more frequently reversed (Possebom, 2023). On the other hand, female judges may exert more effort (informational account) than male judges when analyzing domestic violence cases and, consequently, their rulings might have a smaller probability of being reversed. Importantly, in our empirical context, we find that these two forces balance each other and the Trial Judge's gender has no impact on the outcome of domestic violence cases in the Appeals Court. For this reason, we conclude that the gender conviction-rate gap does not entail a higher cost in terms of Courts' resources.

The impact of stricter female judges on recidivism by defendants in domestic violence cases is also theoretically ambiguous. On the one hand, the positive gender conviction-rate gap may lead to female judges facing more recidivism because stricter judges tend to convict the type of defendants whose recidivism increases when punished (Acerenza et al., 2024). On the other hand, female judges may be able to influence defendants' behaviors in other ways, such as directly signaling that violence against women is not tolerated (representational account). If the typical defendant in a domestic violence case is more likely to recidivate in a similar offense than to commit other types of crime, female judges may lead to lower recidivism. Importantly, in our empirical context, we find that these two forces balance each other and the Trial Judge's gender has no impact on recidivism. For this reason, we conclude that the gender conviction-rate gap has no negative consequences in terms of future criminal behavior by defendants.

This article contributes to the literature on judicial biases. Harris and Sen (2019) reviews this literature, focusing on ideology, race and gender gaps. Within the studies focusing on in-group bias, many authors find evidence of judges making different decisions when they share a group identity with defendants or plaintiffs (Kruttschnitt and Savolainen, 2009; Boyd et al., 2010; Shayo and Zussman, 2011, 2017; Bielen and Grajzl, 2020; Cai et al., 2022; Chen et al., 2022). Similarly to our work, Kruttschnitt and Savolainen (2009), Boyd et al. (2010), Bielen and Grajzl (2020), and Chen et al. (2022) analyze in-group gender bias in cases of uncommon crime types (rape and human trafficking) in Finland, federal appellate cases in the U.S., civil cases in Belgium, and in cases with female offenders in Kenya, respectively. Our work differs from theirs because we focus on a single common crime type (domestic violence) whose victims are mostly women. Furthermore, while Boyd et al. (2010) finds that in-group bias in federal appellate cases in the U.S. is consistent with the informational account only, we find that in-group bias in criminal cases in Brazil is consistent with the representational and informational accounts.

More importantly and differently from the previous studies, we not only document in-group bias by female judges in domestic violence cases but also analyze the consequences of this phenomenon beyond the Trial Court. We investigate the effects of the Trial Judge's gender on the case's outcomes at the Appeals Court and on the defendant's future criminal behavior. In particular, we find that the gender conviction-rate gap has no negative consequences in terms of the use of Courts' resources nor in terms of the defendants' future criminal behavior.

The paper is organized as follows. Section 2 describes our dataset, while Section 3 explains our empirical strategy. Moreover, Section 4 documents the gender conviction-rate gap, while Section 5 discusses its driving forces. Section 6 analyzes the consequences of the gender conviction-rate gap beyond the Trial Court. Section 7 concludes.

2 Data Description and Summary Statistics

To analyze the judicial gender decision gap in domestic violence cases, we use data on criminal cases' characteristics and the judges analyzing those cases. Section 2.1 summarizes our criminal case dataset, while Section 2.2 describes our dataset of judges deciding these cases. Lastly, Section 2.3 explains which judges are included in our sample and presents summary statistics.

2.1 Data on Criminal Cases

We collect data from all criminal cases brought to the Justice Court System in the State of São Paulo, Brazil, between January 4$^{\text{th}}$, 2010, and December 3$^{\text{rd}}$, 2019. This dataset includes the date each case was assigned to a courtroom, the court district where the crime took place, the judge assigned to the case, the full text of the sentence, a list of the main events of the case, the primary and secondary crime types associated to the case, information on whether the case was analyzed by the Appeals Court, and the full text of the Appeals Court's ruling.

The data also contain two pieces of information that are crucial for the construction of our sample, which we discuss in detail in Section 2.3. The first is a dummy indicating if the case was randomly assigned to a judge within a court district and point in time.[2] The second is the name of the courtroom, where different courtrooms may correspond to certain legal specializations. We use this information to make sure that domestic violence cases are indeed randomly assigned within a district and not systematically assigned to (or reassigned to) judges who do not preside over cases of other types of crimes.

We use the list of the case's main events to measure two key case characteristics. First, we measure whether the defendant was caught in the act of committing the crime (*in flagrante delicto*), which is considered strong evidence against the defendant. Second, we observe whether the sheriff temporarily detained the defendant in jail before the case went to the district attorney's office and was assigned a judge. Sheriffs, who must have a Law degree in Brazil, may temporarily detain a defendant if they believe that the defendant poses a significant threat to society.

We use the full-sentence text to construct our main outcome. To do this, we adapt the classification algorithm developed by Possebom (2023) so that it classifies domestic violence case outcomes in a more precise manner.[3] Differently from Possebom (2023), our main outcome focuses exclusively on conviction decisions. This variable equals one when the judge convicts at least one of the defendants listed in the case of at least one of the charges on the case. As a consequence of conviction, the defendant receives some type of punishment (e.g., incarceration, fine, or community service) and a criminal record.

We also collect information on three types of outcomes that may happen after the Trial Judge's ruling. First, we measure whether the case was analyzed by the Appeals Court. This event happens if either the defense attorney or the district attorney disagrees with the Trial Judge's ruling and appeals to a panel of three senior judges. Second, using the classification

[2]Alternatively, some cases are connected to other crimes and are assigned to the judge who was assigned to the previous criminal case.

[3]See Appendix A for an overview of our data-construction process.

algorithm proposed by Possebom (2023), we analyze the full text of the Appeals Court's ruling to measure whether the Trial Judge's ruling was reversed or not.[4] Third, we observe whether the defendant recidivated after their final sentence's date and the type of the new criminal offense. To measure recidivism, we check whether the defendant's name appears in any criminal case within two years after the final sentence date.[5]

Furthermore, the district attorney's office defines each case's primary and secondary crime type prior to judge assignment. We construct three samples based on this classification.

First, our 'domestic violence' sample contains all cases with (i) domestic violence as the main subject or (ii) physical assault as the primary type and domestic violence as the secondary type. The classification "domestic violence" is based on the Maria da Penha Law (Law n. 11,340/2006) and increases the punishment severity of physical assault crimes associated with domestic violence.

Second, our 'physical assault' sample contains cases whose primary type is physical assault. These cases are associated with other forms of assault, such as street fights, bar fights, or football hooliganism. We call 'other physical assault' the physical assault cases that do not involve domestic violence.

Third, our 'misdemeanors' sample contains cases that carry a maximum incarceration sentence of less than four years. This threshold is relevant because, according to Brazilian Law, being convicted of these crime types leads to a criminal record but not to incarceration: sentences must be converted into fines or community service. As discussed by Possebom (2023) and Acerenza et al. (2024), most misdemeanor crimes in Brazil are theft offenses. We call 'other misdemeanors' the misdemeanor offenses unrelated to domestic violence.

Note that the sample of domestic violence cases, other physical assault cases, and other misdemeanor cases are mutually exclusive. We present summary statistics for each of these sub-samples in Section 2.3.

[4] See Appendix A for an overview of our data-construction process.

[5] See Appendix A for an overview of our data-construction process.

2.2 Judges' Career and Characteristics

We collect information on judges' careers from two public sources of information: the annual seniority list published in the São Paulo Justice Diary (2007 to 2019) and the judges' monthly productivity spreadsheet from the São Paulo Court of Justice website (2011 to 2019).

The seniority list ranks all active judges in the court system as of December 31st of each year by career stage and seniority. The dataset includes their starting date as magistrates and the date they were promoted to their current career stage. We combine these pooled seniority lists into a dataset on judges' careers following Laneuville (2024).[6] The judge-level dataset includes the date they entered the justice system, the date of every promotion they have received since December 31st, 2007, and, if they entered the justice system before 2007, the date they reached the career stage they were in at the end of 2007. We link this dataset to the criminal cases dataset by matching the judge's name.

The productivity data from the São Paulo Court of Justice website contains records on the total amount of work done by each judge in any given month and court district from January 2011 to December 2019.[7] We use this information to keep track of which judges are active in each court district-month pair.[8]

2.3 Sample Definition and Summary Statistics

Our final dataset is at the criminal-case level. It is crucial for our identification strategy that the gender of the judge is randomly assigned and correctly identified for each case. To enforce these two characteristics, we adopt four steps.

First, we restrict the dataset to cases that were randomly allocated to judges. By Brazilian Law, a case can be allocated to judges by either connected distribution or free distribution. In the first procedure, the crime under consideration is connected to an existing criminal process.

[6] This dataset allows us to track all variations of judges' names over time. Names may vary over time due to marriage, divorce, or data entry errors.

[7] For example, it measures the number of hearings or disposed cases per month for each judge in each court district.

[8] Months are the natural unit of time for keeping track of activity and work volume because salaries in Brazil are paid monthly. Furthermore, most hiring, reallocation, and termination takes effect at the beginning of each month.

Here, the current case is then attached to the previous process and analyzed by the judge associated with the older case.[9] In the second procedure, the crime under consideration is not connected to an existing criminal process. The new case is then allocated to a judge working in the court district where the offense occurred. If there is more than one judge working in that court district, Brazilian Law dictates that the case be randomly allocated to any judge in the court district. To ensure that the cases in our sample are randomly allocated to judges (and therefore uncorrelated with judge gender), we focus on cases that are allocated according to the free distribution system.

Second, since we focus on domestic violence cases, we must guarantee that this type of case is randomly allocated to a judge. This assumption would be violated if a specific judge or courtroom within the court specialized in domestic violence cases. If there is such a specialized courtroom, then domestic violence cases bypass the free distribution system and are allocated to the specialized judge. For this reason, we restrict the sample to district-quarter pairs without a courtroom specializing in domestic violence.

Third, since we compare decisions made by female and male judges, we require the judge's gender to be randomly allocated to cases.[10] In other words, the *ex-ante* probability of a female judge being assigned to each case must be different from zero or one, which would not be the case in courts without judges of both genders. To enforce this restriction, we limit the sample to cases from court district-month pairs with at least one male and one female active judge in the productivity dataset.

Last, we need to ensure that the assigned judge's gender is determined correctly. To do this, we focus on cases allocated to judges between January 2011 and December 2019 and use our productivity dataset to remove all observations in which the assigned judge was not active at the court district and month of the assignment.[11]

[9] For two simple examples, consider a case where a murderer killed someone and then stole a car to flee the police or a case where a drug trafficker was also caught laundering money.

[10] We measure judges' gender using a two-step procedure. First, we apply the classification algorithm proposed by the *R package* `genderBR`, which uses the share of men and women with each name in the 2010 Brazilian Census. Since a few names are not commonly associated with a specific gender, we visit the São Paulo Court System's website, check the photos of the remaining judges and manually classify their gender. The same classification procedure was used by Laneuville (2024).

[11] Cases are randomly allocated to courtrooms instead of directly to individual judges. If a court district

Using this final sample, we compute summary statistics for cases' outcomes and crime types. Table 1 displays cases' outcomes in the columns and crime types in the rows. It also reports the share of female judges analyzing each type of crime. Domestic violence cases are more frequently convicted than other physical assault cases but are less frequently convicted than other misdemeanor cases. We also note that the share of female judges is similar across crime types and that there are 270 female judges and 425 male judges in our entire sample.

Table 1: Cases' Outcomes and Share of Female Judges by Crime Type in %

Crime Type	Conviction	Female Judge	Obs
DV	32.5	32.1	4,358
Other Mis	36.9	36.4	39,815
Other PA	27.2	38.1	1,648

Notes: The rows of this table indicate the different samples used in this paper: domestic violence cases (DV), misdemeanor cases unrelated to domestic violence (Other Mis), and physical assault cases that are not domestic violence (Other PA). The first column indicates the frequency of cases whose final ruling is conviction. The conviction dummy equals one when at least one defendant is convicted in at least one charge. The second column shows the share of female judges in each sample, while the last column shows the size of each sample.

3 Empirical Strategy and Validity Tests

In this section, we explain the empirical strategy used to measure the gender conviction-rate gap and indirectly test the strategy's validity.

3.1 Empirical Strategy

We measure the gender conviction-rate gap for domestic violence cases by comparing the conviction probability for similar cases allocated to female and male judges. We do this by running a linear regression model on the sample of domestic violence crimes:

$$\text{Convicted}_{it} = \beta \cdot \text{Female Judge}_{it} + \eta \cdot X_{it} + \text{Court District}_i \times \text{Quarter}_t + \varepsilon_{it} \tag{1}$$

is understaffed at the time of case assignment and there is no presiding judge in the assigned courtroom, the system may record the case as being attributed to the next person who takes over the courtroom, which may take place at a later point in time.

where Convicted_{it} indicates whether the cases resulted in a conviction, Female Judge_{it} indicates whether the case was assigned to a female judge, X_{it} are covariates including pre-determined case characteristics (an indicator for whether the defendant was already in jail and an indicator for whether the police caught them committing the crime) and pre-determined judge characteristics (career stage, total experience, and experience in their current career stage as of the day the case was assigned), and "$\text{Court District}_i \times \text{Quarter}_t$" is a full set of court district-quarter pair fixed effects.

Our parameter of interest, β, measures whether a female judge is more likely to convict than a male judge under the assumption that the judge's gender is independent of the case's potential outcome. This exogeneity assumption is valid according to Brazilian Law, which mandates cases to be randomly allocated to judges within court districts (Section 2.3). We indirectly test the validity of this assumption in Table 2.

Even when our exogeneity assumption is valid, the β parameter in Equation (1) may capture differences between male and female judges that are unrelated to their views on domestic violence. To measure whether this conviction-rate gap is due to different gender perspectives about this type of crime, we compare this domestic violence conviction-rate gap against this gap for similar crime types. Now, our sample consists of domestic violence cases and one type of comparable criminal case. We estimate the following linear regression model:

$$\begin{aligned}\text{Convicted}_{it} = \beta_1 \cdot \text{Female Judge}_{it} &+ \beta_2 \cdot \text{Domestic Violence}_{it} \\ &+ \beta_3 \cdot \text{Female Judge}_{it} \times \text{Domestic Violence}_{it} \\ &+ \eta \cdot X_{it} + \text{Court District}_i \times \text{Quarter}_t + \varepsilon_{it},\end{aligned} \tag{2}$$

where "$\text{Domestic Violence}_{it}$" indicates whether the case is associated with domestic violence. We use two categories of comparable crime types: (i) offenses with similar sentences (misdemeanors) and (ii) offenses of a similar nature (physical assault).

In Equation (2), our parameter of interest is β_3. This specification accounts for the extra strictness and social unacceptability of domestic violence offenses beyond that of crimes of comparable severity (coefficient β_2) and for the possibility that female judges are stricter than male judges in general (coefficient β_1). Consequently, the β_3 coefficient captures the part of the gender conviction-rate gap that is likely due to different gender perspectives about domestic violence cases specifically. For this reason, we interpret this coefficient as capturing a type of in-group bias according to the definition given by Shayo and Zussman (2011). We do so because victims of domestic violence are overwhelmingly female, and members of this group (female judges) may act differently from members of the other group (male judges) when analyzing this type of crime in particular.

This interpretation for the β_3 coefficient is valid if the judge's gender is independent of the case's potential outcome after conditioning on the type of crime. This exogeneity assumption is again valid according to Brazilian Law since cases are randomly allocated to judges within court districts, and crime types are determined by the district attorney prior to judge assignment (Section 2.3). We again test this assumption's validity in Table 2.

3.2 Validity Tests

To test the validity of the identifying assumptions behind Equations (1) and (2), we measure the linear relationship between case characteristics and a dummy variable for an assigned female judge. Table 2 shows these correlations for each sample of crimes. We find that, after including court district-quarter fixed effects, judge gender is uncorrelated with case characteristics such as whether the defendant was in jail before the case was assigned, whether the defendant was caught in the act by the police and whether the crime is a domestic violence offense. These results suggest that our identifying assumptions are valid.

4 Measuring the Gender Conviction-Rate Gap

We measure the gender conviction-rate gap using two empirical strategies. First, we estimate Equation (1) and find that female judges are more likely to convict in domestic violence cases than equally senior men. Second, we estimate Equation (2) and find that the gender conviction-rate gap is small or nonexistent for misdemeanors and non-domestic violence

Table 2: Testing the Identification Assumption: Gender of the Judge and Case Characteristics

	Female Judge (1)	Female Judge (2)	Female Judge (3)
Domestic Violence	0.01	-0.01	
	(0.02)	(0.01)	
Caught red-handed?	0.01	0.00	0.02
	(0.03)	(0.01)	(0.03)
Defendant in Jail	-0.02	-0.01	-0.02
	(0.03)	(0.01)	(0.03)
Sample	Physical Assault	Misdemeanors	Domestic Violence
Quarter x Court District FE	✓	✓	✓
Num. obs.	6,006	44,173	4,358

$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$

Notes: Each cell represents the linear relationship between the gender of the judge randomly assigned to the case and a set of case characteristics. Column (1) shows this relationship for the sample of all physical assault crimes, Column (2) shows this relationship for the sample of misdemeanors (crimes such that the maximum sentence does not exceed 4 years), and Column (3) shows this relationship for the sample of domestic violence crimes. All regressions include quarter-by-court district fixed effects.

physical assaults, whereas the gap becomes large and significant for domestic violence cases.

Table 3 shows the effect of judge gender on the probability of conviction in domestic violence cases according to Equation (1). In the first column, we find that randomly selecting a woman or a man to preside does not significantly impact conviction rates.

However, as shown in Laneuville (2024), the average male judge and the average female judge have different career trajectories. We may therefore be concerned that judge tenure and experience may have an impact on conviction rates. For this reason, Column (2) controls for judge career stage, number of years of experience, and number of years in their current position. Here, we find a large effect of gender on the probability of conviction: a woman judge is 31.2% (10 p.p.) more likely to convict a domestic violence offender than a male judge with similar experience. Column (3) shows that this result is robust to the introduction of other case controls.

Since the results in Table 3 may capture differences between male and female judges beyond their gendered perspectives on domestic violence, we also estimate Equation (2) to compare the gender conviction-rate gap for domestic violence cases against this gap for similar cases.

Table 3: Effect of Judge Gender on the Probability of Conviction in Domestic Violence Cases

	Convicted (1)	Convicted (2)	Convicted (3)
Fem	0.04	0.10^{***}	0.10^{***}
	(0.03)	(0.03)	(0.03)
Male Judge's Conviction Rate	0.32	0.32	0.32
Quarter x Court District FE	✓	✓	✓
Career Controls		✓	✓
Case Controls			✓
Num. obs.	4,358	4,358	4,358

$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$

Notes: This table presents the estimated results of Equation (1). The first column shows the effect of judge gender on the probability of conviction. The second column includes controls for judge characteristics (career stage, number of years of experience as a judge, and number of years of experience in the current career stage). The third column controls for both judge and case characteristics (whether the defendant was caught red-handed and whether the defendant was arrested by the sheriff before the trial). All regressions include quarter-by-court district fixed effects.

Table 4 shows the estimated results associated with Equation (2). The first three columns use the sample of misdemeanors, therefore comparing crimes with similar sentence lengths to those in the domestic violence sub-sample. The last three columns use the sample of physical assault cases, which are similar in nature to domestic violence cases as they entail physical harm to another person.

Table 4 shows that the estimated coefficient on the female dummy (β_1) is small for the misdemeanors sample and close to zero for the physical assault sample. This suggests that the gender conviction-rate gap is small or nonexistent for comparable crimes that are not classified as involving domestic violence.

More importantly, Table 4 shows that the coefficient on the interaction of judge gender and domestic violence (β_3) is statistically significant and positive in all regressions, regardless of included control variables. The effect is particularly strong for the physical assault sample (Columns (4)-(6)). This result suggests that the gender difference in conviction rates verified in Table 3 is neither fully explained by the severity or the nature of domestic violence offenses nor by female judges being stricter than male judges in all rulings.

Consequently, the interaction coefficient captures the part of the gender conviction-rate gap that is specific to domestic violence cases and is likely due to different approaches to domestic violence cases between male and female judges. For this reason, this result suggests the existence of a type of in-group bias according to the definition given by Shayo and Zussman (2011).[12] In the next section, we explore the driving forces behind this in-group bias, while, in Section 6, we investigate its consequences for the cases' later stages in the Justice System and the defendants' future behavior.

[12] In Appendix B, we propose a simple economic model formalizing the definition of in-group bias.

Table 4: Effect of Judge Gender on the Probability of Conviction Comparing Domestic Violence Cases against Similar Crimes

	Sample: Misdemeanors			Sample: Physical Assault		
	Convicted	Convicted	Convicted	Convicted	Convicted	Convicted
	(1)	(2)	(3)	(4)	(5)	(6)
Female	0.02^{**}	0.02^{***}	0.02^{***}	-0.03	-0.00	0.00
	(0.01)	(0.01)	(0.01)	(0.04)	(0.04)	(0.04)
DV	-0.05^{***}	-0.05^{***}	-0.04^{***}	0.06^{***}	0.06^{**}	0.05^{**}
	(0.01)	(0.01)	(0.01)	(0.02)	(0.02)	(0.02)
DV x Female	0.03^{*}	0.04^{*}	0.03^{*}	0.07^{*}	0.09^{**}	0.08^{**}
	(0.02)	(0.02)	(0.02)	(0.04)	(0.04)	(0.04)
Men's CR (Non-DV)	0.38	0.38	0.38	0.30	0.30	0.30
Men's CR (All)	0.37	0.37	0.37	0.31	0.31	0.31
Quarter-District FE	✓	✓	✓	✓	✓	✓
Career Controls		✓	✓		✓	✓
Case Controls			✓			✓
Num. obs.	44,173	44,173	44,173	6,006	6,006	6,006

$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$

Notes: This table presents the estimated results of Equation (2). The first three columns compare differences in male and female judge conviction rates in domestic violence cases to the gender difference in conviction rates for misdemeanors. The last three columns compare differences in male and female judge conviction rates in domestic violence cases to the gender difference in conviction rates for other physical assault crimes. CR stands for Conviction Rate and DV stands for Domestic Violence.

5 Understanding the Gender Conviction-Rate Gap

In this section, we show two important findings about the conviction rate gender gap and its driving forces.

The first finding is suggestive evidence that the gap is explained by judges' identification with their own gender (Section 5.1). We find that female judges are more likely than male judges to convict and simultaneously flag the case as an intimate partner violence offense, while female and male judges are equally likely to convict and not flag the case as an intimate partner violence offense. This result suggests that the representational account (Boyd et al., 2010) is one of the explanatory forces of the in-group bias documented in Table 4 because violence between relationship partners is highly impacted by patriarchal social norms that make gender identity more salient (Heise and Kotsadam, 2015; Tur-Prats, 2019; González and Rodríguez-Planas, 2020).

The second finding, discussed in Section 5.2, is that men and women respond differently to evidence presented in the same case. We show that female judges are more likely than male judges to convict in cases where evidence is more tenuous and subject to interpretation, but not when the evidence is objectively incriminating (i.e., when the defendant was caught red-handed). This result suggests that the informational account (Boyd et al., 2010) is one of the explanatory forces of the in-group bias documented in Table 4.

In Appendix B, we propose a simple threshold-crossing model that formalizes the concepts of the representational and informational accounts.

5.1 Female Identity as a Driver for Larger Conviction Rates

In this section, we show that the gender of the judge matters more in cases where patriarchal social norms are particularly relevant: intimate partner violence. We interpret this finding as indicative that a major reason for the gender conviction-rate gap is that domestic violence cases make gender identity more salient to judges, suggesting that the representational account (Boyd et al., 2010) is one of the explanatory forces of the in-group bias documented in Section 4 because female judges tend to protect member of their own group more intensely than male judges.

We argue that domestic violence cases make gender identity more salient because sexist social norms and gender roles significantly affect the occurrence and punishment of intimate partner violence in multiple contexts. For instance, González and Rodríguez-Planas (2020) show that, among immigrant families, the incidence of intimate partner violence is positively affected by gender norms in the country of origin. Heise and Kotsadam (2015) show that the incidence of intimate partner violence in different countries is strongly associated with patriarchal social norms, such as male authority over women, justifications for wife beating, and laws and practices that disadvantage women in property ownership. Closer to our context, Perova and Reynolds (2017) show that police stations that specialize in domestic violence with specially trained officers significantly reduce the number of deaths of young women by intimate partner violence in metropolitan municipalities in Brazil. In terms of judicial decision-making on domestic violence, the negative stereotypes mentioned by Kafka et al. (2019) and Human Rights Watch (1991) concern female intimate partners.

To show that the gender of the judge matters more for the outcome of intimate partner violence cases than for that of other domestic violence offenses, we take advantage of an important feature of our data. We are able to identify the relationship between the victim and the assailant for all cases that culminate in an active decision by the presiding judge (conviction or acquittal). To do this, we construct an algorithm that defines a case as intimate partner violence if, when writing the court's decision, the judge uses the Portuguese word for wife, girlfriend, ex-wife, ex-girlfriend, or similar words.[13]

We, then, test whether female judges are more likely to make a conviction decision in a domestic violence case and simultaneously flag the case as being related to intimate partner violence after controlling for case characteristics. If there is a gender conviction-rate gap in cases flagged as intimate partner violence but not in cases without such a flag, then a stronger identification with the victim may be driving the gender lenience gap found in Section 4. If

[13] In the domestic violence sample, 57.8% of cases are flagged as violence against partners or ex-partners. The specific words we look for are "namorada," "ex-namorada," "esposa," "ex-esposa," "amásia," "ex-amásia," "companheira," "ex-companheira," "mulher,", "ex-mulher," "coabitação," "união estável," "casamento," "era casado," and "era casada."

the gender conviction-rate gap is present regardless of explicit mentions to intimate partner violence, then there must be other forces driving this lenience gap.

To implement this test, we interact the "intimate partner violence" indicator with the conviction indicator. This interaction term becomes our new outcome variable in Equation (1).

Table 5 shows the results of these regressions. In Column (1), the outcome is a dummy variable indicating whether the judge's ruling mentioned intimate partner violence and reached a conviction decision. In Column (2), the outcome is a dummy variable indicating whether the judge's ruling did not mention intimate partner violence but reached a conviction decision. In Column (3), the outcome is a dummy variable indicating whether the case resulted in a conviction, replicating the results from Table 3 for convenience.

We find that female judges are 37.5% (9 p.p.) more likely than male judges to convict the defendant and simultaneously flag the case as intimate partner violence (Column (1)). However, female judges are as likely as male judges to convict the defendant while not flagging the case as intimate partner violence (Column (2)). These results suggest that the gender conviction-rate gap is higher in cases where a female judge may identify more strongly with the victim compared to a male judge.

These findings are consistent with existing evidence that, in many contexts, judges make different decisions when they share a group identity with defendants or plaintiffs (Kruttschnitt and Savolainen, 2009; Shayo and Zussman, 2011, 2017; Bielen and Grajzl, 2020; Cai et al., 2022; Chen et al., 2022). Many of these papers show that the differences intensify as group identity becomes more salient. This intensification may result from of a heightened opposition between two or more social groups ("us" versus "them"). For example, Shayo and Zussman (2011) show that Jewish Israeli judges in small claim courts favor Jewish litigants more intensely when they are more exposed to conflict-related terrorism. Closer to our case, Cai et al. (2022) show that judges analyzing divorce cases in China favor plaintiffs of their own gender more in areas where social norms are more conservative. In our case, we show that the gender conviction-rate gap is much more intense in cases where the form of domestic violence

Table 5: Effect of Judge Gender on the Probability of Conviction and the Salience of Intimate Partner Violence

	Convicted and Partner (1)	Convicted and Not Partner (2)	Convicted (3)
Female	0.09***	0.01	0.10***
	(0.03)	(0.01)	(0.03)
Men's Outcome Average	0.24	0.07	0.32
Quarter-District FE	✓	✓	✓
Case Controls	✓	✓	✓
Judge Career Controls	✓	✓	✓
Num. obs.	4,358	4,358	4,358

$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$

The regressions in this table refer to outcomes for the sample of domestic violence crimes only. The outcome in Column (1) is whether the case involved intimate partner violence and led to a conviction. The outcome in Column (2) is whether the case did not involve intimate partner violence and led to a conviction. The outcome in Column (3) is whether the case led to a conviction. All regressions include quarter-district fixed effects and a full set of controls for case characteristics and judge career stages.

is more stereotypical.[14] Consequently, the representational account (Boyd et al., 2010) is a plausible driving forcing of the in-group bias documented in Section 4 because female judges tend to protect members of their own group more intensely than male judges.

5.2 Male and Female Judges Analyze Evidence Differently

In this section, we show that female judges are more likely than male judges to convict in domestic violence cases where evidence is more tenuous and subject to interpretation, but not when the evidence is objectively incriminating. We interpret this result as suggesting that the informational account (Boyd et al., 2010) is one of the explanatory forces of the in-group bias documented in Table 4 because female judges process the information contained in domestic violence cases differently from their male peers.

Domestic violence cases require complex evidentiary rulings. As discussed by Beecher-Monas (2001), these cases may involve both victim testimony about the incident(s) and expert witness testimony about battered woman syndrome. Since such testimony is inherently

[14] Note that most domestic violence cases against women are committed by partners and ex-partners (United Nations Office on Drugs and Crime, 2023).

subjective, female and male judges may interpret testimony evidence differently.

Our dataset allows us to investigate whether the gender conviction-rate gap is stronger in cases where the evidence set is less clearly incriminating. For each case, we observe whether the defendant was arrested because they were caught in the act of committing the crime (i.e., *in flagrante delicto*). These cases have more objective evidence, often including testimony from police officers who directly witnessed the offense.

If the gender conviction-rate gap is stronger in cases where the defendant was not caught red-handed, then the gender leniency gap may be because female and male judges interpret more subjective pieces of evidence differently (informational account). If the gender conviction-rate gap is similar across domestic violence cases regardless of the type of evidence presented, then there must be other forces driving this lenience gap.

To measure the gender conviction-rate gap for cases with different evidence types, we re-estimate Equations (1) and (2) using two sub-samples. The first regression restricts our sample to cases where the defendant was caught red-handed, while the second regression restricts our sample to cases where the defendant was not caught *in flagrante delicto*.

Table 6 reports regression results for the two different samples. The odd columns include cases where the defendant was caught red-handed, while the even columns include cases where this objective evidence was not present. Columns (1) and (2) report the results associated with Equation (1), while Columns (3)-(6) report the results associated with the interaction model described in Equation (2).

We find that a gender conviction-rate gap exists only in those cases where the defendant was not caught red-handed. We only find significant results for the female coefficient (β) in Column (2) and the interaction coefficient (β_3) in Column (6). These results suggest that female and male judges interpret less objective evidence differently, while they interpret objective evidence (cases flagged as *in flagrante delicto*) in similar ways. These different interpretations of more subjective cases are consistent with the informational account being one possible driving force of the in-group bias documented in Section 4.

Table 6: Effect of Judge Gender on the Probability of Conviction conditional on the Type of Evidence

	Sample: Domestic Violence		Sample: Misdemeanors		Sample: Physical Assault	
In flagrante delicto	Yes	No	Yes	No	Yes	No
	(1)	(2)	(3)	(4)	(5)	(6)
Fem	0.04	0.13^{***}	0.02	0.03^{**}	0.05	0.01
	(0.06)	(0.04)	(0.01)	(0.01)	(0.13)	(0.04)
DV			-0.14^{***}	-0.01	0.03	0.07^{**}
			(0.02)	(0.01)	(0.08)	(0.03)
DV x Fem			0.05	0.03	-0.02	0.12^{**}
			(0.03)	(0.02)	(0.13)	(0.05)
Men's CR (Non-DV)	-	-	0.49	0.33	0.40	0.28
Men's CR (All)	0.35	0.30	0.47	0.32	0.36	0.30
Quarter-District FE	✓	✓	✓	✓	✓	✓
Judge Career Controls	✓	✓	✓	✓	✓	✓
Num. obs.	1,337	3,021	14,391	29,782	1,552	4,454

$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$

Notes: This table compares the determinants of the probability of conviction. Cases are split into a sub-sample in which the defendant was caught red-handed (odd columns) and a sample in which the defendant was not caught red-handed (even columns), requiring the justice system to rely on less objective evidence. The first two columns are constructed using the sample of domestic violence cases only. Columns (3) and (4) use the sample of misdemeanors. Columns (5) and (6) use the sample of physical assault crimes. CR stands for Conviction Rate and DV stands for Domestic Violence. All regressions include quarter-district fixed effects and a full set of controls for judge career stages.

6 Consequences of the Gender Conviction-Rate Gap

A gender conviction-rate gap for domestic violence cases, as found in Section 4, may have consequences for the cases' later stages in the Justice System and the defendants' future behavior. In particular, defense attorneys, district attorneys, Appeals Court judges, and defendants may change their behavior when facing a female judge in a domestic violence case.

In Section 6.1, we investigate whether stricter female judges may lead to more appeals and ruling reversals in domestic violence cases. Although we may expect positive gender appeal-rate and reversal-rate gaps due to the fact that stricter judges are more frequently reversed by the Appeals Court (Possebom, 2023, Figure 1), we find that female judges' rulings in domestic violence cases have similar outcomes in the Appeals Court when compared to male judges' rulings.

In Section 6.2, we investigate whether stricter female judges may lead to more recidivism by defendants in domestic violence cases. Since stricter female judges might be more likely to convict the type of defendants whose recidivism increases when punished (Acerenza et al., 2024), we may expect a positive gender recidivism gap. Despite this expectation, we find that female and male judges face similar recidivism probabilities in their domestic violence cases.

6.1 Consequences of the Gender Conviction-Rate Gap at the Appeals Court

A gender conviction-rate gap for domestic violence cases may impact how the defense attorneys, district attorneys and Appeals Court judges treat domestic violence cases after the Trial Judge has issued a ruling. In this section, we investigate whether these agents change their behavior when facing a female judge in a domestic violence case.

On the one hand, the phenomenon of stricter female judges for domestic violence cases may lead to more appeals and reversals of trial judges' rulings. This positive gender gap may be expected for two reasons.

First, Possebom (2023, Figure 1 and Table J.1) finds that rulings by strict judges are more frequently reversed by the Appeals Court than rulings by more lenient judges, partially because defense attorneys are more likely to appeal than district attorneys. Combining this result with our finding that female judges are stricter than male judges when analyzing domestic violence cases, we may expect a positive gender gap in the probability of domestic violence cases being analyzed by the Appeals Court and in the probability of domestic violence rulings being reversed by the Appeals Court.

Second, Laneuville (2024, Table 3) states that 93.55% of Appeals Court judges who were active between 2007 and 2022 were men. If we assume that judges of both genders behave similarly at the Trial Court and the Appeals Court, our results suggest that male Appeals Court judges may be more lenient when analyzing domestic violence cases and reverse Trial Judges' conviction rulings more frequently. Consequently, we may expect a positive gender gap in the probability of domestic violence rulings being reversed by the Appeals Court.

On the other hand, the informational account for the gender conviction-rate gap in domestic violence cases (Section 5.2) may lead to a smaller probability of female judges having their rulings reversed by the Appeals Court. According to the informational account of the in-group bias, female judges analyze domestic violence evidence in a different way, possibly implying that they write more complete or careful rulings. When we combine this possibility with the fact that the Appeals Court focuses more on the ruling's form than the case's substance, we may expect a negative gender gap in the probability of domestic violence rulings being reversed by the Appeals Court.

We investigate which of these two forces dominates by analyzing if there is a gender gap in the probability of domestic violence cases being analyzed by the Appeals Court and in the probability of domestic violence rulings being reversed by the Appeals Court. To do so, we re-estimate Equations (1) and (2) using two new outcome variables. First, our left-hand side variable is an indicator that equals 1 if there is an appeal process associated with the case and 0 otherwise. Second, our outcome variable is an indicator that equals 1 if the Appeals Court reversed the Trial Judge's ruling and 0 otherwise. To avoid sample selection issues, note that the "zero" group includes all cases that were not analyzed by the Appeals Court.

Table 7 shows the estimated coefficients of Equations (1) and (2) when we use our appeal and reversal indicators as outcome variables. The left-hand side variable is the appeal indicator in the odd columns, while it is the reversal indicator in the even columns. The first two columns are constructed using the sample of domestic violence cases only (Equation (1)). Columns (3) and (4) use the sample of misdemeanors (Equation (2)), while Columns (5) and (6) use the sample of physical assault crimes (Equation (2)). All regressions include quarter-district fixed effects and a full set of controls for case characteristics and judge career stages.

There are three results worth highlighting. First, the sign and significance of the female coefficient in Columns (3) and (5) show that female judges are slightly more likely to be appealed in misdemeanor and physical assault cases, respectively. This finding is not surprising because female judges are stricter than male judges (Column (3) in Table 4), and defense attorneys appeal more frequently than district attorneys (Possebom, 2023, Table J.1). Second, because female coefficients are statistically null in Columns (2), (4) and (6), we find that, even though female trial judges are more frequently appealed, there is no gender difference in the reversal probability. Third and more importantly, the coefficients associated with the consequences of the in-group bias ("female" in Columns (1) and (2) and "DV x Female" in Columns (3)-(6)) are not statistically significant.[15] The null result in the "female" coefficient in Columns (1) and (2) means that female and male trial judges are equally likely to be appealed and reversed in domestic violence cases. The "DV x Female" coefficient in Columns (3)-(6) means that the gender gap in the probability of being appealed and reversed is not significantly more intense in domestic violence cases than in other comparable cases.

These null results suggest that the forces driving the gender appeal-rate and reversal-rate gaps upwards and downwards compensate each other. In other words, even though female judges are stricter than male judges when analyzing domestic violence cases, their rulings are complete and careful enough so that they do not generate more reversal decisions by the Appeals Court.

[15] Remember that the in-group bias captures the part of the gender conviction-rate gap that is specific to domestic violence cases.

Table 7: Effect of Judge Gender on the Probability of a Case Being Analyzed by the Appeals Court and of a Trial Sentence Being Reversed by the Appeals Court

	Sample: DV only		Sample: Misdemeanors		Sample: Physical Assault	
	Appealed	Reversed	Appealed	Reversed	Appealed	Reversed
	(1)	(2)	(3)	(4)	(5)	(6)
Female	0.04	0.01	0.02^{**}	0.00	0.05^{*}	0.01
	(0.03)	(0.01)	(0.005)	(0.004)	(0.03)	(0.01)
DV			-0.09^{***}	-0.03^{***}	0.04^{*}	0.01
			(0.01)	(0.01)	(0.02)	(0.01)
DV x Female			0.01	0.01	-0.01	0.00
			(0.02)	(0.01)	(0.04)	(0.01)
Men's Average (Non-DV)			0.37	0.05	0.17	0.03
Men's Average (All)	0.29	0.03	0.37	0.05	0.22	0.03
Quarter-District FE	✓	✓	✓	✓	✓	✓
Career Controls	✓	✓	✓	✓	✓	✓
Case Controls	✓	✓	✓	✓	✓	✓
Num. obs.	3,668	3,668	46,251	46,251	5,868	5,868

$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$

Notes: This table shows the results of Equations (1) and (2) using two new outcome variables. In the odd columns, the outcome variable is binary and indicates whether the case was analyzed by the Appeals Court, while, in the even columns, the outcome variable is binary and indicates whether the trial judge's sentence was reversed by the Appeals Court. The first two columns are constructed using the sample of domestic violence cases (Equation (1)). Columns (3) and (4) use the sample of misdemeanors (Equation (2)). Columns (5) and (6) use the sample of physical assault crimes (Equation (2)). DV stands for Domestic Violence. All regressions include quarter-district fixed effects and a full set of controls for case characteristics and judge career stages.

6.2 Consequences of the Gender Conviction-Rate Gap for the Defendants' Recidivism Behavior

A gender conviction-rate gap for domestic violence cases may impact how defendants behave after their sentences. In this section, we investigate whether these agents change their behavior when facing a female judge in a domestic violence case.

On the one hand, the phenomenon of stricter female judges for domestic violence cases may increase the probability of recidivism. This positive gender gap may be expected for two reasons.

First, Acerenza et al. (2024) analyzes misdemeanor cases in São Paulo during our sample period and finds that punishment increases two-year recidivism for agents who would be punished only by strict judges (type A defendant) while it decreases two-year recidivism for agents who would be punished by most judges, including the lenient judges (type B defendant). In our context, we find that female judges are stricter than male judges (Table 4, Column (3)), especially for domestic violence cases (Table 4, Columns (3) and (6)). Moreover, if we assume that female judges are stricter because they have a harsher bar for conviction (instead of having more information about domestic violence cases), then we expect female judges to convict more agents who are more likely to recidivate when punished (type A defendant) while being similarly likely to convict type B defendants. Consequently, we may expect a positive gender gap in the probability of a defendant recidivating.

Second, domestic violence victims may be encouraged by conviction decisions to accuse their aggressors if they recidivate by committing a new domestic violence offense. If this is the case, we may expect a positive gender gap in the recidivism probability because female judges are stricter than male judges when analyzing domestic violence cases (Section 4). To avoid this mechanical increase in the recidivism probability, we also analyze the effect of a female judge on the probability of committing a new crime that is not a domestic violence offense.

On the other hand, the representational account for the gender conviction-rate gap in domestic violence cases (Section 5.1) may lead to a smaller probability of recidivism. For example, female judges may be able to directly influence the behavior of the defendants by better signaling that violence against women is not tolerated. If the typical defendant in a domestic violence case is more likely to recidivate in a similar offense than to commit other types of crime, female judges may lead to lower recidivism. Consequently, we may expect a negative gender gap in the recidivism probability.

We investigate which of these two forces dominates by analyzing if there is a gender gap in the probability of a defendant recidivating within two years of the case's final sentence. To do so, we re-estimate Equations (1) and (2) using two new outcome variables. First, our left-hand side variable is an indicator that equals 1 if the defendant commits any new criminal offense within two years of their final sentence and 0 otherwise. Second, our outcome variable is an indicator that equals 1 if the defendant commits any new non-domestic violence offense within two years of their final sentence and 0 otherwise.

Table 8 shows the estimated coefficients of Equations (1) and (2) when we use these two measures of recidivism as outcome variables. The left-hand side variable is the any-recidivism indicator in the odd columns, while it is the non-domestic violence recidivism indicator in the even columns. The first two columns are constructed using the sample of domestic violence cases only (Equation (1)). Columns (3) and (4) use the sample of misdemeanors (Equation (2)), while Columns (5) and (6) use the sample of physical assault crimes (Equation (2)). All regressions include quarter-district fixed effects and a full set of controls for case characteristics and judge career stages.

Note that the coefficients associated with the consequences of the in-group bias ("female" in Columns (1) and (2) and "DV x Female" in Columns (3)-(6)) are not statistically significant.[16] We also highlight that the relevant point estimates in Columns (1)-(4) are either small or negative.

These results suggest that the forces driving the gender recidivism gap upwards and downwards compensate each other. In other words, even though female judges are stricter than male judges when analyzing domestic violence cases, their stricter rulings are not criminogenic.

[16]Remember that the in-group bias captures the part of the gender conviction-rate gap that is specific to domestic violence cases.

Table 8: Effect of Judge Gender on the Probability of a Defendant Recidivating

	Sample: DV only		Sample: Misdemeanors		Sample: Physical Assault	
	Any	Excluding DV	Any	Excluding DV	Any	Excluding DV
	(1)	(2)	(3)	(4)	(5)	(6)
Female	0.01	-0.02	0.01	0.00	-0.02	-0.04
	(0.04)	(0.04)	(0.01)	(0.01)	(0.05)	(0.05)
DV			-0.02	-0.06^{***}	0.02	-0.01
			(0.02)	(0.01)	(0.04)	(0.04)
DV x Female			0.01	0.00	0.04	0.03
			(0.02)	(0.02)	(0.06)	(0.05)
Men's Average (Non-DV)			0.35	0.35	0.25	0.24
Men's Average (All)	0.32	0.28	0.35	0.34	0.30	0.27
Quarter-District FE	✓	✓	✓	✓	✓	✓
Career Controls	✓	✓	✓	✓	✓	✓
Case Controls	✓	✓	✓	✓	✓	✓
Num. obs.	2,665	2,665	21,319	21,319	3,467	3,467

$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$

Notes: This table shows the results of Equations (1) and (2) using two new outcome variables. In the odd columns, the outcome variable is binary and indicates whether the case's defendant was prosecuted for any crime after their final sentence's date, while, in the even columns, the outcome variable is binary and indicates whether the case's defendant was prosecuted for any crime that is not a domestic violence offense. The first two columns are constructed using the sample of domestic violence cases (Equation (1)). Columns (3) and (4) use the sample of misdemeanors (Equation (2)). Columns (5) and (6) use the sample of physical assault crimes (Equation (2)). DV stands for Domestic Violence. All regressions include quarter-district fixed effects and a full set of controls for case characteristics and judge career stages.

7 Conclusion

In this paper, we provide evidence that domestic violence cases assigned to female judges are significantly more likely to result in conviction than those assigned to male judges with similar career paths. This pattern is, to a large extent, due to the reasons that are specific to domestic violence cases. The gender difference in conviction rates is stronger for cases involving domestic violence than for crimes that are similar in nature (physical assault) and sentence length (misdemeanors), suggesting the existence of a type of in-group bias as defined by Shayo and Zussman (2011).

Furthermore, we find suggestive evidence for two major drivers of the gender conviction-rate gap in domestic violence. The first driver is that female judges are stricter than men in cases of intimate partner violence but equally strict in cases of other types of domestic violence. We claim this finding shows the relevance of identification with one's own gender, as the perception of intimate partner violence is highly affected by patriarchal social norms, while other forms of violence are not affected by gender stereotypes. This result is consistent with the representational account as defined by Boyd et al. (2010).

The second driver comes from gender differences in the way judges analyze evidence. There is no gender gap when the defendant is caught red-handed by the police (clear and objective piece of evidence), but the gap is very strong when there is more room for subjectivity in the analysis. This result is consistent with the informational account as defined by Boyd et al. (2010).

Lastly, we find that the gender conviction-rate gap has no impact on the probability of appeals, ruling reversals or recidivism for female judges in domestic violence cases. Consequently, we conclude that there is no evidence that the gender conviction-rate gap affects the use of Courts' resources or the defendants' future criminal behavior.

References

Acerenza, S., V. Possebom, and P. H. C. Sant'Anna (2024, January). Was Javert right to be suspicious? Unpacking treatment effect heterogeneity of alternative sentences on time-to-recidivism in Brazil. arXiv:2311.13969 [econ].

Beecher-Monas, E. (2001). Domestic Violence: Competing Conceptions of Equality in the Law of Evidence Symposium on Integrating Responses to Domestic Violence - Dedication. *Loyola Law Review 47*(1), 81–136.

Bielen, S. and P. Grajzl (2020, September). Gender-based judicial ingroup bias in sex crime sentencing: Evidence from Belgium. *International Journal of Law, Crime and Justice 62*, 100394. tex.lccn: Gender.

Boyd, C. L., L. Epstein, and A. D. Martin (2010). Untangling the Causal Effects of Sex on Judging. *American Journal of Political Science 54*(2), 389–411.

Cai, X., P. Li, Y. Lu, and H. Song (2022, September). Do Judges Exhibit Gender Bias? Evidence from the Universe of Divorce Cases in China.

Chen, D. L., J. Graham, M. Ramos Maqueda, and S. Singh (2022, March). Do Judges Favor Their Own Ethnicity and Gender? *World Bank Symposium*. Publisher: Washington, DC: World Bank tex.lccn: Gender/Ethnic.

Ferraz, C. and L. Schiavon (2022, May). Crime, Punishment, and Prevention: The Effect of a Legal Reform on Violence Against Women.

González, L. and N. Rodríguez-Planas (2020, October). Gender norms and intimate partner violence. *Journal of Economic Behavior & Organization 178*, 223–248.

Harris, A. P. and M. Sen (2019). Bias and Judging. *Annual Review of Political Science 22*(1), 241–259. _eprint: https://doi.org/10.1146/annurev-polisci-051617-090650.

Heise, L. L. and A. Kotsadam (2015, June). Cross-national and multilevel correlates of partner violence: an analysis of data from population-based surveys. *The Lancet Global Health 3*(6), e332–e340. Publisher: Elsevier.

Human Rights Watch (1991, October). Criminal Injustice: Violence Against Women in Brazil. Technical report.

Kafka, J. M., K. E. Moracco, C. Barrington, and A. L. Mortazavi (2019, July). Judging Domestic Violence From the Bench: A Narrative Analysis of Judicial Anecdotes About Domestic Violence Protective Order Cases. *Qualitative Health Research 29*(8), 1132–1144.

Kruttschnitt, C. and J. Savolainen (2009, May). Ages of Chivalry, Places of Paternalism: Gender and Criminal Sentencing in Finland. *European Journal of Criminology 6*(3), 225–247. Publisher: SAGE Publications.

Laneuville, H. (2024, June). Breaking the Glass Ceiling with a Gavel? Gender Promotion Gap in the Brazilian Judiciary.

Perova, E. and S. A. Reynolds (2017, February). Women's police stations and intimate partner violence: Evidence from Brazil. *Social Science & Medicine 174*, 188–196.

Possebom, V. (2023, September). Crime and Mismeasured Punishment: Marginal Treatment Effect with Misclassification. *The Review of Economics and Statistics*, 1–42.

Shayo, M. and A. Zussman (2011, August). Judicial Ingroup Bias in the Shadow of Terrorism. *The Quarterly Journal of Economics 126*(3), 1447–1484.

Shayo, M. and A. Zussman (2017, October). Conflict and the Persistence of Ethnic Bias. *American Economic Journal: Applied Economics 9*(4), 137–165.

Tur-Prats, A. (2019, 12). Family Types and Intimate Partner Violence: A Historical Perspective. *The Review of Economics and Statistics 101*(5), 878–891.

United Nations Office on Drugs and Crime (2023, December). *Gender-related Killings of Women and Girls (Femicide/feminicide): Global Estimates of Female Intimate Partner/family-related Homicides in 2022*. United Nations.

Supporting Information

(Online Appendix)

A Additional Details about the Criminal Cases Data

In this appendix, we explain how we construct our dataset about criminal cases in São Paulo. We build upon the dataset collected by Possebom (2023), modifying it to allow for a deeper analysis of domestic violence cases. For additional legal definitions, check Appendix I.2 by Possebom (2023).

The final dataset was created from three initial datasets.

1. CPOPG ("Consulta de Processos de Primeiro Grau"): It contains information about all criminal cases in the Justice Court System in the State of São Paulo (TJ-SP) between 2010 and 2019.

2. CJPG ("Consulta de Julgados de Primeiro Grau"): It contains information about the last decision made by a trial judge in all criminal cases in TJ-SP between 2010 and 2019.

3. CPOSG ("Consulta de Processos de Segundo Grau"): It contains information about all appealing criminal cases in TJ-SP between 2010 and 2019.

Starting from the CPOPG dataset, we implement the following steps.

1. We only keep cases that are currently in the Appeals Court, closed, or whose status is empty. Those cases are already associated with a trial judge's sentence.

2. We only keep cases that aim to analyze whether a defendant is guilty or not.

3. We only keep cases that were randomly assigned to trial judges.

4. We only keep cases whose starting date is after January 1$^{\text{st}}$, 2010.

We then merged it with the CJPG dataset using cases' `id` codes.

After this step, we randomly select 65 cases per year (2010-2019) for manual classification. From this sample, we select 35 cases each year to form our training sample, ensuring that at least 10 of them are physical assault cases. Moreover, we select 10 cases each year to form our overall validation sample, and 10 domestic violence cases and 10 extra physical assault cases to form our crime-type-specific validation samples.

We manually classify these cases into six categories: "defendant died during the trial", "case was expired", "defendant is guilty", "defendant accepted a non-prosecution agreement" ("transação penal" in Portuguese), "case was dismissed" ("processo suspenso" in Portuguese) and "defendant was acquitted". Since some sentences are missing or incomplete, we are able to manually classify only 611 sentences.

Now, we use those 611 manually classified cases to train a classification algorithm. To do so, we divide them into a training sample (315 cases), an overall validation sample (109 sentences), a domestic violence-specific validation sample (99 sentences), and a physical assault-specific validation sample (88 sentences).

First, we design an algorithm to identify which defendants died during the trial. To do so, we check whether the sentence contains any reference to the first paragraph of Article 107 from the Brazilian Criminal Code. This deterministic algorithm classifies cases into the category "defendant died during the trial" almost perfectly. It misclassified only one sentence that had two defendants and only one of them died before the judge made a decision.

Second, we design an algorithm to identify which cases expired. To do so, we check whether the sentence contains any reference to the fourth paragraph of Article 107 from the Brazilian Criminal Code. This deterministic algorithm classifies cases into the category "expired case" almost perfectly. It misclassified only four sentences that included multiple crime types in one single case.

Third, we design an algorithm to identify which cases were dismissed. To do so, we check whether the sentence contains any reference to Article 89 in Law n. 9099/95. This deterministic algorithm correctly classifies 98% of the cases into the category "case was dismissed".

Fourth, we design an algorithm to identify which defendants accepted a non-prosecution agreement. To do so, we check whether the sentence contains any expression connected to a non-prosecution agreement. This deterministic algorithm correctly classified almost all the cases into the category "defendant accepted a non-prosecution agreement", making only five mistakes.

Finally, we design an algorithm to classify the remaining cases into two categories: "defendant is guilty" and "defendant was acquitted". To do so, we define a bag of words that were found to be strong signals of acquittal and guilt when manually classifying the cases in our samples.[17] We then count how many times each one of those expressions appears in each sentence, and we normalize those counts to be between 0 and 1.

Using the normalized counts, we train an L1-Regularized Logistic Regression using our training sample. We then validate this algorithm using our three validation samples and find that it correctly classifies 97.9% of the overall cases, 97.9% of the domestic violence cases, and 95.3% of the physical assault cases. Given this high success rate, we use the L1-Regularized Logistic Regression algorithm to classify which trial judge rulings concluded that the "defendant is guilty" in our full sample.

Having designed the above algorithm, we use it to define whether the trial judge reached a conviction ruling, which is our main outcome variable. First, we find which defendants died during their trials and drop them from our sample. We then use the second and third algorithms to define which cases were dismissed and which cases are associated with a non-prosecution agreement. Moreover, we use the trained L1-regularized Logistic Regression algorithm to classify the remaining cases into the categories "defendant is guilty" and "defendant was acquitted". In the end, our main outcome value equals 1 if the case was classified into the category "defendant is guilty" and equals 0 otherwise.

We also use the CPOPG dataset to measure recidivism (Section 6.2). We focus on whether the defendant recidivated within 2 years of their case's final sentence. Following Possebom (2023), a defendant i in a case j recidivated ($Y_{ij} = 1$) if and only if defendant i's full name appears in a case $\bar{j}$ whose starting date is within 2 years after case j's final sentence's date.

[17] Possebom (2023, Appendix I.2) lists the entire bag of words.

Importantly, case $\bar{j}$ can be about any type of crime, including very severe crimes, while case j has to be about domestic violence, misdemeanor, or physical assault offenses. To match defendants' names across cases, we use the Jaro–Winkler similarity metric and define a match if the similarity between full names in two different cases is greater than or equal to 0.95.[18]

Moreover, in Table 8's even columns, we exclude from the recidivism definition any case that is a domestic violence case. Specifically, we observe the offense type of case $\bar{j}$ and do not count it as a recidivism event if it is a domestic violence offense.

Importantly, in Section 6.2 only, we delete the case-defendant pairs whose cases started in 2018 and 2019 because their recidivism variable is not properly defined due to right-censoring.

Lastly, we use the CPOSG dataset to measure whether a case was analyzed by the Appeals Court and whether the Appeals Court's ruling reverses the Trial Judge's ruling.

We start by merging the CJPG dataset with the CPOSG dataset using each case's `id` code. When merging these datasets, we create an indicator variable that denotes which cases were analyzed by the Appeals Court, i.e., which cases were matched. This is the outcome variable in Table 6.1's odd columns.

We, then, randomly select 50 cases per year for manual classification (2010-2019) and divide them into three categories: "cases that went to the Appeals Court, but were immediately returned due to bureaucratic errors", "cases whose trial judge's rulings were affirmed" and "cases whose trial judge's rulings were reversed".

Now, we use those 500 manually classified cases to train a classification algorithm. To do so, we divide them into a training sample (300 cases) and a validation sample (200 sentences).

First, we design an algorithm to identify which cases went to the Appeals Court but were immediately returned. To do so, we simply check whether the Appeals Court's decision is missing in our dataset.

Finally, we design an algorithm to classify the non-empty cases into two categories: "cases whose trial judge's sentences were affirmed" and "cases whose trial judge's sentences were

[18] Abramitzky et al. (2019) match full names in historical Censuses in the U.S. and Norway. They define a match between two individuals if the Jaro–Winkler similarity between their names is greater than or equal to 0.90 and if their dates of birth match exactly. Since we do not observe defendants' dates of birth, we adopt a stricter Jaro-Winkler similarity threshold to define a match in our dataset.

reversed". To do so, we define a bag of words that were found to be strong signals of sentence reversal when we manually classified the cases in our sample.[19] We, then, count how many times each one of those expressions appears in each sentence, and we normalize those counts to be between 0 and 1.

Using the normalized counts, we train an L1-Regularized Logistic Regression using our training sample. We then validate this algorithm using our validation sample and find that it correctly classifies 96.2% of the cases. Given this high success rate, we use the L1-Regularized Logistic Regression algorithm to classify which trial judge's rulings were reversed in our full sample. This is the outcome variable in Table 6.1's even columns.

[19] Possebom (2023, Appendix I.2) lists the entire bag of words.

B Modelling the Gender Conviction-Rate Gap

In this appendix, we propose a simple economic model to formalize the concepts of gender conviction-rate gap and in-group bias (Shayo and Zussman, 2011), and their driving forces, the representational and informational accounts (Boyd et al., 2010). First, to formally define the first two concepts, we adapt the potential outcome model (Rubin, 1974) to a scenario where cases of different types of crime may be treated with a female judge instead of a male judge. Second, to formally define the representational and informational accounts of the in-group bias, we adapt the threshold-crossing model (Heckman and Vytlacil, 2005) to a context where the latent heterogeneity is associated with the evidence of a criminal case and the propensity score function is associated with the judge's conviction criteria.

We start by looking at criminal cases as our observational unit. Each case is associated with a type of crime: $d = 1$ indicates that it is a domestic violence offense, and $d = 0$ indicates that it is either a misdemeanor offense or another type of physical assault case. Moreover, each case may receive two randomly allocated treatments: $f = 1$ indicates that the case was assigned to a female judge and $f = 0$ indicates that the case was assigned to a male judge. Finally, each case has a potential conviction decision, $Y(f, d)$, which depends on the gender of the presiding judge and on the type of crime.[20]

We define the gender conviction-rate gap as

$$\mathbb{E}\left[Y(1,1) - Y(0,1)\right] \neq 0, \tag{B.1}$$

i.e., the probability that a domestic violence case ends in a conviction when presided by a female judge is different from the probability that a domestic violence case ends in a conviction when presided by a male judge.

[20] Note that, in our context, sex is a manipulable variable because the unit of observation is the case instead of the judge. Moreover, type of crime is a pre-determined covariate, because it is chosen by the prosecutor before the presiding judge is assigned.

We also define the in-group bias as

$$\begin{cases} \mathbb{E}\left[Y\left(1,1\right)-Y\left(0,1\right)\right] & > & 0 \\ \mathbb{E}\left[Y\left(1,1\right)-Y\left(0,1\right)\right] & > & \mathbb{E}\left[Y\left(1,0\right)-Y\left(0,0\right)\right] \end{cases}, \tag{B.2}$$

i.e., the gender conviction-rate gap for domestic violence cases (i) is positive, and (ii) is larger than the gender conviction-rate gap for other types of crime.

Next, we formally define the representational and informational account using our empirical context and a threshold-crossing model. In any criminal case, evidence is produced against the defendant. In our model, the judge sets a conviction criterion and interprets the evidence. Then, the judge convicts the defendant if the judge interprets the evidence as being sufficient to do so.

The evidence interpretation function, $E\colon \left\{0,1\right\}^3 \times \mathbb{R} \rightarrow \mathbb{R}$, depends on the judge's gender (a female indicator variable, F), the type of crime (a domestic violence indicator variable, D), whether there is objective and strong evidence against the defendant (an indicator variable R that is equal to one if the defendant was caught red-handed when committing the crime), and all other types of evidence produced by the police and the prosecution, V. Hence, $E\left(F,D,R,V\right)$ is a random variable capturing the amount of evidence against the defendant as interpreted by the judge assigned to the case.

The judge's conviction criterion function, $P\colon \left\{0,1\right\}^3 \rightarrow \mathbb{R}$, depends on the judge's gender, the type of crime and whether the crime is associated with intimate partner violence (an indicator variable W that is equal to one if the victim is the wife or intimate partner of the defendant). Hence, $P\left(F,D,W\right)$ is a random variable capturing the conviction criterion of the judged assigned to the case.

The judge convicts the defendant if the judge interprets the evidence as being sufficient to do so, i.e.,

$$Y = \mathbf{1}\left\{P\left(F,D,W\right) \leq E\left(F,D,R,V\right)\right\}, \tag{B.3}$$

where Y denotes whether the defendant is convicted or not.

Note that Equation (B.3) imposes two exclusion restrictions. While the relationship status between the victim and the defendant — W —- only enters the judge's conviction criterion function, being caught red-handed — R — only enters the evidence interpretation function. The first exclusion restriction is imposed because the Criminal Process Code does not include this variable as valid evidence against the defendant, but the Criminal Law Code includes this variable as an aggravating factor when setting the sentence. The second exclusion restriction is imposed because the Criminal Process Code considers *in flagrante delicto* as a piece of evidence against the defendant but not as an aggravating factor.

Using this model, we can formally define the representational account driving the in-group bias as

$$P(1,1,1) - P(0,1,1) < P(1,1,0) - P(0,1,0) \leq 0, \tag{B.4}$$

i.e., female judges may set stricter conviction criteria than male judges for domestic violence cases and female judges are even stricter when the case is associated with intimate partner violence.

We can also formally define the informational account driving the in-group bias as

$$E(1,1,0,v) - E(0,1,0,v) > E(1,1,1,v) - E(0,1,1,v) \geq 0, \tag{B.5}$$

for any v in the support of V. When these inequalities hold, the female judge interprets the evidence more harshly than a male judge when the defendant is caught red-handed and the female judge interprets the evidence even more harshly when the defendant is not caught red-handed.

In the main text, we use data from criminal offenses in the State of Sao Paulo, Brazil, to test the existence of each one of the four concepts defined above. In Table 3, we show that the gender conviction-rate gap (Equation (B.1)) exists. Then, in Table 4, we show that in-group bias (Equation (B.2)) is present in our application. Finally, Table 5 shows that our data is consistent with the representational account (Equation (B.4) while Table 6 provides evidence that our data is consistent with the informational account (Equation (B.5).

References

Abramitzky, R., L. Boustan, and K. Eriksson (2019). To the New World and Back Again: Return Migrants in the Age of Mass Migration. *ILR Review 72*(2), 300–322.

Boyd, C. L., L. Epstein, and A. D. Martin (2010). Untangling the Causal Effects of Sex on Judging. *American Journal of Political Science 54*(2), 389–411.

Heckman, J. J. and E. Vytlacil (2005). Structural equations, treatment effects, and econometric policy evaluation1. *Econometrica 73*(3), 669–738.

Possebom, V. (2023, September). Crime and Mismeasured Punishment: Marginal Treatment Effect with Misclassification. *The Review of Economics and Statistics*, 1–42.

Rubin, D. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology 66*(5), 688–701.

Shayo, M. and A. Zussman (2011, August). Judicial Ingroup Bias in the Shadow of Terrorism. *The Quarterly Journal of Economics 126*(3), 1447–1484.