

Yan, S., & Lao, J. (in press). Sex disparities in sentencing and judges' beliefs: A vignette approach. *Victims & Offenders*.

This is a post-peer-review, pre-copyediting version of a paper accepted for publication at *Victims & Offenders*. This paper is not the version of record and may not exactly replicate the authoritative version. The final article is available, upon publication, at a DOI address to be provided by the publisher.

## Sex Disparities in Sentencing and Judges' Beliefs: A Vignette Approach

Shi Yan<sup>1</sup> and Jiaqi Lao<sup>2,3</sup>

<sup>1</sup> School of Criminology and Criminal Justice, Arizona State University

<sup>2</sup> College for Criminal Law Science, Beijing Normal University

<sup>3</sup> Law School, Beijing Normal University

### Author Note

Shi Yan  <https://orcid.org/0000-0002-6484-7687>

We have no known conflict of interest to disclose. We thank the anonymous reviewers for their comments, and Siyu Liu and Becky Leung for their input. Early results of the study were presented at a seminar at Southern Illinois University, and we thank the comments from the audience as well. All errors remain our own.

Correspondence concerning this article should be addressed to Shi Yan, 411 N Central Ave, Ste 600, Phoenix, AZ 85004. Email: [shiyang@asu.edu](mailto:shiyang@asu.edu)

## Sex Disparities in Sentencing and Judges' Beliefs: A Vignette Approach

### Abstract

The predominant theories on courtroom decision-making explain extralegal disparities with the courtroom actors' use of stereotypes. We conducted a vignette study on a sample of judges at the Chinese National Judges College, manipulated the sex of the defendant in each vignette, then asked the subjects for the recommended sentence. The survey also contained a series of questions on the beliefs about the causes of crime, the patterns of criminal behaviors, and the effectiveness of punishment. We found that the judges recommended significantly less harsh sentences for the female homicide defendant, but recommended significantly harsher sentences for both female defrauding and drug trafficking defendants. We also found little evidence that the perception and belief variables were confounders behind the observed sex disparities.

Keywords: sentencing, women as offenders, courts

## Sex Disparities in Sentencing and Judges' Beliefs: A Vignette Approach

### Introduction

More than 40 years ago, Armstrong (1977) argued that women were “‘protected’ but unequal” under the law. Archaic and grotesque some readers today may find (but see Chesney-Lind, 2020), she condemned the state statutes that explicitly prescribed different sentences for men and women who committed a similar crime. This disparate treatment between men and women was not limited within the United States, but pervaded many parts of the world (e.g., Pittaway & Bartolomei, 2001; Rapaport, 1991). While many of the laws allegedly “protected” women, Armstrong found the sentencing disparity between sexes “the most blatant evidence” for the denial of women’s fundamental rights (p. 120).

Over the decades, both scholars and policymakers view sentence disparities associated with defendants’ demographic characteristics as unwarranted and concerning (Baumer, 2013; Daly & Tonry, 1997). Despite the consensual goal of sentencing uniformity, empirical findings on sex disparities, as well as the underlying mechanism, were mixed. There have been two competing theories on how the state treats female defendants: “the evil woman” perspective and the chivalry perspective. Many studies found that female defendants were treated less harshly than comparable male defendants, while some found either no significant disparities or harsher penalties for women (for reviews, see Bontrager et al., 2013; Daly & Bordt, 1995; Gaub & Holtfreter, 2015). Historically, most of the foundational work on sex disparities in sentencing was based on the analysis of court data obtained from the United States. Researchers from around the world are now introducing their perspectives and findings on sex disparities (e.g., Drápal,

2020; Liu et al., 2018; Lu et al., 2013; Pina-Sánchez & Harris, 2020).

Nevertheless, this growing literature is still primarily based on official records, which typically offers a good sample size but only contains a limited number of case and judge characteristics.

The present study further contributes to the increasing intellectual diversity of the literature with a unique research setting. We surveyed over 270 Chinese judges using an experimental vignette design, manipulating the sex of the defendants in four different scenarios. We also collected information on the judges' beliefs and perceptions of crimes and the criminal justice system, a critical missing piece in almost all administrative datasets available today. Empirical research on sentencing in China is burgeoning recently, thanks to the nationwide publication of court decisions online (Liu et al., 2018; Wu, 2020a, 2020b). Yet to our awareness, efforts of primary data collection from judges—regardless of the country—remain scant (Ulmer, 2012). Our research not only provides an international perspective on the issue of sex disparities, but also contributes to the fundamental understanding of sentencing as a human decision.

The paper proceeds as follows. First, we will summarize the theories and empirical findings on sex disparities in sentencing, with a balanced narrative on both the American and the Chinese contexts. Then we will discuss the role of judicial beliefs and perceptions in shaping the sentence. Next, we will present our data, analysis, and results. We conclude our paper by discussing the substantive significance and implications of our findings.

**Sex Disparities in Sentencing: How Much Overlap is There between the U.S. and China?**

### ***Feminist Theories and the Criminal Justice Response***

The foundation of modern feminist theories on crime and criminal justice originated from the United States (Chesney-Lind, 2006, 2020; Daly & Chesney-Lind, 1988; MacKinnon, 1983). The theories and social movements emerged in an era when “girls and women in conflict with the law were overlooked or excluded in mainstream works while demonized, masculinized, and sexualized in the marginalized literature that brooded on their venality” (Chesney-Lind, 2006, p. 7). Researchers called for the theorization of gender and the shaping of gender through social institutions.<sup>1</sup> The inquiries into sex disparities in case outcomes also came to the attention of researchers (Spohn et al., 1985; Steffensmeier, 1980).

In the court and sentencing context, two competing theoretical themes emerged (Daly & Tonry, 1997; Etienne, 2010; Farnworth & Teske, 1995; Spohn & Beichner, 2000). One of the perspectives claimed that chivalry pervaded in the courtrooms. The courtroom actors saw the necessity to protect female defendants from a patriarchal standpoint, and were therefore likely to issue less harsh sentences. To the contrary, another perspective argued that the courtroom actors saw women who commit crimes as violators of the assumed gender roles and norms (often known as “the evil woman”), and therefore more culpable and dangerous than their male counterparts. In addition to the chivalry-“evil women”

---

<sup>1</sup> Throughout the paper, we use “sex” to refer to the biological sex of people, and “gender” to refer to the cultural norms and expectations associated with people’s biological sex. In the overwhelming majority of criminal justice studies, researchers can only measure the sex of people (and therefore sex disparities in case outcomes), but many theories (including the main ones used in this study) are based on *gender* roles and norms. The two concepts are especially prone to confusion in the criminal justice context, and researchers sometimes use them interchangeably. While an in-depth differentiation between the two concepts is beyond the scope of our study, we aim to use the two terms in a precise manner.

debate, more general theories on sentencing also viewed gender as an essential extralegal factor associated with the sentence. Under both the bounded rationality (Albonetti, 1991) and the focal concerns (Steffensmeier et al., 1998) perspectives, prosecutors and judges may also associate the sex of the defendant with stereotypes in the assessment of risk and blameworthiness.<sup>2</sup>

Among all the different angles and standpoints, researchers have attempted to reach a more flexible framework, known as the selective chivalry perspective (Farnworth & Teske, 1995; Spohn & Beichner, 2000). The perspective contended that the presence and direction of sex disparities in the sentence are not universal, but instead depends on the specific case circumstances.<sup>3</sup> Both the criminal justice system and the society would treat women harsher when they violated gender norms, and less harshly when there were more understandable causes behind their behaviors. For example, studies on the media found that women were more likely to receive negative coverage when they commit crimes against children, presumably due to a graver violation of their perceived role as caretakers (Christensen, 2018; Grabe et al., 2006). However, actors in the criminal justice system were more likely to show sympathy to women who commit violent actions against an abusive male partner (Miller, 2001, p. 1339, emphasis added):

---

<sup>2</sup> Although both theoretical frameworks were much more commonly associated with racial disparities in sentencing (Ulmer, 2012), many studies on sex disparities or the race-sex interaction effects also utilized them, since the blameworthiness and risk components are compatible with the themes of gender-centered theories (see Bontrager et al., 2013). For simplicity, the paper focuses on the gender-centered theories (i.e., chivalry and “the evil woman”) whenever possible, and only applies the more general perspectives when necessary.

<sup>3</sup> Some studies used the term “selective chivalry” to denote a perspective closer to “the evil woman” hypothesis (Embry & Lyons, 2012; Rodriguez et al., 2006). Throughout the paper, we stick to the original scopes of the selective chivalry perspective.

Beth cut her husband's throat so badly... he almost died. He was constantly abusing her... and at the time of the stabbing, she said he was beating the crap out of her... *That's what she felt she had to do to get out of the situation.* -Probation Officer #1

Most theoretical accounts of gender in the Chinese context emerged from historical and comparative analyses, with a focus on the interplay between the social and cultural factors that moved gender dynamics in different directions (for comprehensive examinations, see Croll, 1978; Mann, 2011; Santos & Harrell, 2017). On the one hand, patriarchal ideology and social structure had been profoundly influential for centuries. On the other hand, the socialist movement's embrace of egalitarian ideals, the influence of the West accompanying the "reform and opening-up," and the dissolution of the traditional family structure have all contributed to higher levels of gender equality (Lu et al., 2013; Yan, 2006). The interaction clash between the traditional and recent forces is also highly visible in the realm of sex/gender and crimes (for a recent review, see Shen & Winlow, 2014). For example, researchers have long attributed sexual exploitation and domestic violence against women in China to the highly patriarchal traditional family values (Gil & Anderson, 1999; Ruskola, 1994). Yet recent studies on women's crime and victimization in China also had more contemporary scopes: urbanization, migration worker families, and so on (Liu, 2012; Niu & Laidler, 2015).

Theoretical development on sex disparities in today's Chinese courts is still mostly embryonic. Most, if not nearly all, empirical studies followed the themes of chivalry versus "the evil woman" (Liang et al., 2009; Lu et al., 2013; Wei & Xiong, 2020). Despite the overlap in the framework, researchers realized that gender

norms in both the U.S. and Chinese contexts were dynamic and evolving. Therefore, the range and degree of chivalry (or the absence thereof) varied, sometimes dramatically. Early in the 20<sup>th</sup> century, Chinese statutes saw women who assaulted or murdered the husband or male family members as committed an aggravated crime (Ruskola, 1994). Even in the early 2000s, judges across China meted out was a wide range of sentences for women who killed their husbands after domestic violence victimization—*from three years on probation to death without suspension* (Xing, 2013). The nature of gender roles in shaping the sentence appears to be intricate.

### ***Empirical Supports of the Two Perspectives***

Despite the hopeful claims that “the differential treatment of women in sentencing... is a thing of the past” (Chesney-Lind & Pasko, 2013, p. 142) and “‘chivalry’ is alive and well” (Spohn, 1999, p. 392), the empirical literature suggested a more mixed pattern from large datasets on case processing. Indeed, literature reviews and meta-analyses across three decades found overall less harsh treatments of female defendants (Bontrager et al., 2013; Daly & Bordt, 1995; Gaub & Holtfreter, 2015; Zatz, 2000), and many recent studies continued to claim so (Cassidy & Rydberg, 2020; Fridel, 2019; Nowacki, 2019; Pina-Sánchez & Harris, 2020). However, the finding of a lower harshness level for women is not universal and monotonic across the board. Studies also found that the presence and direction of sex disparities depended on the race of the defendant (Goulette, 2021), the type of crime (Bond & Jeffries, 2014; Rodriguez et al., 2006), or criminal records (Tillyer et al., 2015). Overall, there appears to be a mix of

evidence supporting not universal chivalry, but selective chivalry—that the presence and direction of sex disparities depend on other case-level factors.

The same dependency and inconsistency are also true among studies using Chinese case data. A series of studies by Liang and colleagues found that while female defendants generally received less harsh outcomes, the magnitude and statistical significance of the leniency depended on the type and seriousness of the crime (Liang et al., 2009; Lu et al., 2013; Xiong et al., 2018). More recently, studies using larger and more representative case samples from China found diminished to no disparities associated with the sex of either the defendant or the judge (Wei & Xiong, 2020; Xia et al., 2019). Similar to studies using U.S. data, the support of the theories in the Chinese context is also mixed. Less harsh punishment for female defendants was visible in some crime types and circumstances, but not others.

A more complicated issue is that the gender norms—the focal point to understand sex disparities in the sentence—are not always clearly visible and straightforward. In some cases like personal violence (especially when the victim is a family member or a child), the aggravating (such as predatory acts) and mitigating (such as domestic violence victimization) elements associated with the defendant's sex are relatively unambiguous. However, for many other types of crimes, such as theft, drug trafficking, and drunk driving, there is less societal agreement on what the gender norms are, or whether a widely accepted norm exists at all. Although the selective chivalry hypothesis is flexible, its explanative and predictive function still rests on gender norms. Given the lack of consensus

for many crime types, the present study also serves as an exploratory effort to connect the selective chivalry framework with more crime types.

### **The Role of Judicial Beliefs and Its Testing**

The current literature has highlighted the role of gender norms and expectations, as we summarized above. Theories of punishment have also indicated that the belief in the root cause of crimes can affect the decision of individual cases or the general attitude penal policies. Specifically, individuals who emphasize free will would likely conclude that crimes are the voluntary choice of the offender, and the offender should be responsible for the losses and damages. Therefore, they are less likely to tolerate crimes and more likely to support decisions based on deterrence and retribution (Finckenauer, 1978; Frase, 1997). To the contrary, those who perceive social inequality and economic adversities as the root cause of crimes are more likely to support rehabilitation-based criminal justice policies (Carroll et al., 1987). Empirical work found that judges and mock judges were more likely to mete out lenient sentences when they believed that crimes were beyond the defendants' rationality, and harsher penalties when they thought that crimes were products of free will (Aspinwall et al., 2012; Bumby & Maddox, 1999; Hill, 1998; Martin et al., 2017; Shariff et al., 2014). In a recent vignette-based experimental study of real judges, Aspinwall et al. found that providing expert testimony detailing the defendant's biological development of psychopathy led to less harsh recommended sentences. The testimony suggested that the crime was the result of a disease rather than a mere voluntary act.

Despite the findings, two research gaps remain. First, most of the studies examined the direct relationship between those belief variables and the outcomes.

Few have measured the beliefs under the context of extralegal disparities and explored whether judicial beliefs played a confounding role under the observed disparities between sexes. Second, unlike the research on extralegal disparities in case outcomes, the literature on Chinese judges' beliefs is especially barren. Most seminal research on sentencing disparities in China was based on analyses of case files—either datasets compiled by researchers (Bai, 2013; Liang et al., 2009; Lu et al., 2013) or the more recent data released by the Chinese Supreme People's Court on their affiliated website (Ahl & Sprick, 2018; Liebman, 2015; Wu, 2020a, 2020b). While those studies gave us a broad view of disparities in Chinese courts, administrative datasets are, by definition, incompatible with inquiries into the belief variables. Presumably due to accessibility issues, data collection from Chinese judges is rare, and most of such studies utilized qualitative techniques—interviews and observations (Clarke, 2003; Kinkel & Hurst, 2015; Zheng et al., 2017). We are unaware of recently published studies that surveyed Chinese judges on their beliefs. More generally, relative to analyses of administrative court data, experimental studies on courts remains scarce even in the United States (Aspinwall et al., 2012; Bushway et al., 2014; Fishbane et al., 2020; Wilford et al., in press). The need for judges as experiment subjects is only stronger in the Chinese context.

Although there is evidence on the association between judges' beliefs and case outcomes, the relationship between the beliefs and extralegal disparities remains unclear both theoretically and empirically. Just like the pattern of gender norms in the less stereotypical crimes committed by women, the current status of literature remains largely exploratory. Based on the assumption of the theories on

sex disparities, one potential way to validate the theories is to measure judges' views on gender norms and stereotyping directly. However, a validity concern behind the measures is that views associated with gender stereotypes may be perceived as unpopular (Hayes et al., 2001; Ho & McLeod, 2008), and especially so among judges given their role in the society. As a result, our study employed a set of less direct measures of judges' perceptions and beliefs about crimes, offenders, and the criminal justice system (detailed in Method section below). While we are more confident in the truthfulness of the responses, a potential downside of our approach is the inability to establish a causal model explaining sex disparities directly.

### **Study Settings and Research Hypotheses**

China is a valuable study site for our research question for both theoretical and practical reasons. While gender norms are everchanging around the world (Bustelo, 2009; Kaufman & Williams, 2013), their rapid changes in China have resulted in a unique set of dynamics (Liu, 2012; Niu & Laidler, 2015). In addition to their role as professionals, judges are also members of their communities, which may have equipped them with different views on gender norms and roles. More importantly, the Supreme People's Court has issued sets of binding regulations aiming at enhancing the consistency among decisions, reducing unwarranted disparities, and curbing case-level discretion (Ahl, 2014; Roberts & Pei, 2016). While researchers have studied the level of extralegal disparities extensively through case decision databases, one advantage of vignette studies is it allows for the manipulation of the focal variable (sex of the defendant in our study) while holding all other case characteristics constant, which reduces the noise level in

theory testing. Practically, in most countries around the world (and especially in China, Ahl & Sprick, 2018; Clarke, 2003), primary data collection from real judges is much more costly and logistically challenging than from college students or nationally representative adult samples. While it inevitably limits our sample size, judge subjects do have the expertise in sentencing, making the findings having higher levels of both internal and external validity than those obtained from general public samples.

Given the theories on selective chivalry and the empirical findings summarized above, we hypothesize that our analysis of sex disparities will also reveal a mixed pattern. While there will be evidence of favorable treatment of female defendants (chivalry), there will also be evidence suggesting a null effect or unfavorable treatment of women. Instead of universal leniency on women, we expect the direction and significance of the disparities to depend on the nature of crimes. With respect to the perception and belief variables, we hypothesize that judges' beliefs of the criminal justice system and the nature of crimes will also play a role in shaping the sentence. Specifically, judges who attributed crimes to the free will of defendants were more likely to recommend harsher sentences than those who attributed crimes to factors outside of the defendants' control. We also hypothesize that these beliefs will help to explain the sex disparities we observe.

## **Method**

### **Data Collection and Analytic Sample**

We conducted data collection at the Chinese National Judges College, located in Beijing, China. The National Judge College offers both initial appointment training and continuing training for judges. We administered the

study in paper-and-pen format and gave out informed consent before distributing the survey. We distributed the study to the entirety of judges in two different training sessions in late 2016 and early 2017 (170 judges each, 340 total), and received 272 completed and valid surveys (response rate = 80.0%). Among the completed surveys, there was no evidence of major missing data problems.<sup>4</sup> Both our *N* and our response rate were comparable to recent vignette and experiment studies of judges conducted in the United States (Aspinwall et al., 2012; Spamann & Klöhn, 2016).

### Instrument and Manipulation

The original study instrument was entirely written in Chinese. We have attached the full instrument in both Chinese and English (translated by ourselves) as Supplemental Files to the manuscript for peer review and archival purposes.<sup>5</sup> The first part, identical for all subjects, inquired about the judges' demographics and professional background. The second part, also identical for all subjects, measured the judges' perceptions and beliefs. We asked one question about their *belief on the responsible party of crimes* (the offender, the society, or half and

---

<sup>4</sup> Out of all independent variables used in the analysis, none was missing for more than 12 subjects (4.41% of completed surveys). Three of the questions on the comparison between offenders had a "not sure/does not know" option. Instead of dropping the subjects who selected those options (between 6.6% and 11.2% of completed surveys), we kept them as a separate category. Out of the 272 judges, 17 (6.25%) had a missing recommended sentence type or length for one or more vignettes. We imputed the missing values with the median length for each of the crime types (120 months for homicide, 72 months for robbery, 36 months for defrauding, and 84 months for drug trafficking). We also tried to leave the fields as missing and used listwise deletion, and the substantial findings were very similar.

<sup>5</sup> The original instrument also contained a belief question about defendants' recidivism patterns, and sentence recommendations if the defendants had prior records. We did not utilize those questions in the current study because we believed they fall under a different scope, and the more appropriate way to explore these issues is in a different study.

half), which corresponds to the prior literature on individual free will versus societal causes of crimes. We asked three questions on the judges' *belief on the differences between people who committed crimes and law-abiding people, between people who committed different types of crimes, and between male and female offenders* (very different, somewhat different, almost the same, not sure/does not know). All those questions measured the use of perceptual shorthands or stereotypes. We also asked two questions on the judges' beliefs on the prison's effectiveness of the punishment and the prison's effectiveness of rehabilitation (4-point Likert scale, 1 = very effective, 4 = not effective at all).

The third and last part of the instrument contained four case vignettes: a homicide case, a robbery case, a defrauding case, and a drug trafficking case. Ideally, we would fully manipulate all possible defendant sex combinations across the vignettes and create 16 different versions (2 by 2 by 2 by 2). However, with that many different versions, we would have needed an unrealistically large sample size to maintain the necessary statistical power. Instead, we designed two versions of the instrument, Version M and Version F. All defendants in Version M vignettes were men, while all defendants in Version F vignettes were women. Except for the homicide victim (the opposite-sex spouse of the defendant), the sex of all victims and co-defendants were identical between the vignette versions (see Supplemental Files for full vignettes with case details). After each vignette, we asked the judges the type of sentence they would have meted out if they heard the case (fixed-term imprisonment, life imprisonment, death with a suspension, and death without a suspension), as well as the length for fixed-term imprisonment (in months). We did not provide judges with the option of a probation-only

sentence, because real-life defendants charged with any of the crimes depicted would only have a minimal chance to receive probation.

For reasons we were unable to capture in the survey process, the response rate was much higher for Version F ( $n = 155$ , 91.2% response rate) than for Version M ( $n = 117$ , 68.9% response rate). A balance test found that the observed characteristics of judges differed between the two conditions, suggesting non-random assignment. Therefore, we controlled for all observed judge characteristics in our analyses, and caution our readers to be aware of potential omitted variable bias in our findings.

### **List of Variables and Analytic Strategy**

The dependent variable in the study was the sentence length, measured in months. We coded life imprisonment as 240 months, the death penalty with a suspension as 360 months, and the death penalty without a suspension as 600 months, in light of the common practice in China.<sup>6</sup> In addition to the main analysis using this coding scheme, we also conducted two sensitivity checks regarding the dependent variables. For the robbery, drug trafficking, and defrauding scenarios, at least 96% of judges recommended fixed-term imprisonment, and the coding scheme of life and death sentences would only have a minimal impact on the

---

<sup>6</sup> The vast majority of defendants receiving life imprisonment do get released after serving approximately 20 years through commutation. Most defendants receiving a suspended death sentence will first have their sentence reduced to life imprisonment, then eventually get released in a similar fashion. While China did enact a life without parole article in 2015 (before our survey), it is unlikely to be applicable to any of the defendants depicted in the vignettes. For additional information on life sentences in China, see Smith and Jiang (2019); for additional information on the suspended death penalty, see Trevaskes (2013). Throughout the nearly 1100 judge-vignette combinations (272 judges \* 4 vignettes), the death penalty without a suspension was only recommended twice, both for the drug trafficking scenario, which suggests its coding should have no impact on three of the four vignettes, and only a minimal impact for drug trafficking.

findings. However, the distributions of the sentence length for all three vignettes were positively skewed (raw skewness between 2.66 and 5.73). We estimated the models using log-transformed dependent variables (skewness between -0.20 and 0.63 after transformation) to address the skewness issue. Although there was no major dependent variable skewness issue with the homicide vignette (raw skewness = 0.48), over a third of judges recommended either life imprisonment or the suspended death penalty, making the findings potentially more sensitive to the modeling choices. We conducted a Heckman two-step correction analysis to account for the potential selection effect into fixed-term imprisonment.

Our primary independent variable was the sex of the defendant. The list of independent variables also includes the judges' age, sex, educational level (1 = postgraduate degree, 0 = bachelor's degree), the level of the court they worked for at the time of survey (1 = primary, 2 = intermediate, 3 = higher), and the set of belief and perception variables detailed above. We included all these variables in the main analysis to maximize the amount of information. However, given our small sample size, we are also aware of the necessity of model parsimony—that having too many independent variables might have taken away the degrees of freedom. In the last sensitivity test, we collapsed the court level and perception of responsibility variables into dummies, as well as combined the three perceived difference variables and two perceived effectiveness variables into two composite scores by taking the respective average values. We present all sensitivity tests in the Results section after the main analysis.

## **Results**

### **Descriptive Statistics**

In Table 1, we present the descriptive statistics for the sample. Over two-thirds of the judges were male, and a third of them had a graduate degree. On average, the judges were 40.0 years old. Judges from primary and intermediate courts each constituted about 40% of the sample, and slightly over 20% of our sample came from a higher court.

Nearly all judges (98.9%) believed that the offender was at least somewhat responsible for crimes. Specifically, 56.9% of judges believed that both the offender and the society were equally responsible for crimes, and 42.0% believed that the offender was entirely responsible. The perception questions had a very different response pattern. Only 8.9% of the judges considered people who committed crimes to be very different from law-abiding people. However, 27.7% of the judges considered people who committed different types of crimes to be very different, and 33.2% considered male and female offenders to be very different.<sup>7</sup> Overall, judges had a moderate view on the effectiveness of prisons. However, the average score for punishment effectiveness was slightly below the midpoint ( $M = 2.3$ ), and the average score for rehabilitation effectiveness was slightly above the midpoint ( $M = 2.7$ ). Given our coding scheme, the scores suggest that judges slightly agreed with the notion that prisons could punish, but slightly disagreed with the notion that prisons could rehabilitate.

For all four vignettes, the modal recommended sentence type was fixed-term imprisonment. For the robbery, defrauding, and drug trafficking vignettes, over 96% of judges recommended fixed-term imprisonment. There was a larger

---

<sup>7</sup> A pairwise correlation analysis found a low-to-modest level of correlation among the perception and belief variables ( $|r| \leq .45$  for all pairs), suggesting that the variables were capturing different dimensions of judges' beliefs.

disagreement on the sentence for the homicide vignette. Over 60% of judges recommended fixed-term imprisonment, while 28.0% and 7.4% of judges recommended life imprisonment and the death penalty respectively. It is also noteworthy that among judges who recommended the death penalty for the homicide defendant, all recommended a suspension. There was some considerable variation in the recommended length of imprisonment, both among vignettes and among judges within a given vignette. Among the fixed-term imprisonment terms (i.e., converted life/death sentences and imputed sentences not included), the homicide defendants received the longest average sentence recommendation ( $M = 114.1$ ,  $SD = 45.8$ ), followed by the robbery ( $M = 83.7$ ,  $SD = 38.2$ ) and drug trafficking defendants ( $M = 81.7$ ,  $SD = 31.3$ ). The defrauding defendant received the shortest average recommended sentence ( $M = 44.5$ ,  $SD = 21.9$ ). It is noteworthy that for all four vignettes, the range of recommended sentence lengths was ten years or longer, thus suggesting a visible level of disagreement among the judges.

## Regression Analysis

### *Main Findings*

Table 2 presents the findings of the regression analysis, with the length of the sentence as the dependent variable (this time including the converted length for life/death sentences and imputed length for missing cases). In three out of the four vignettes, there was a statistically significant sex disparity in the judges' recommended sentence. However, the direction of the disparities was not consistent across the models. For the homicide vignette, the male defendant received an average recommended sentence 43.8 months *longer* than the female

defendant did. For the defrauding and drug trafficking vignettes, the male defendants received an average recommended sentence 10.8 months and 20.9 months *shorter* than the female defendants did, respectively.

Although we included a set of judge characteristics for our models, these variables generally did not appear to be confounding the sex disparities we found. In three out of the four vignettes (except for drug trafficking), male judges gave out harsher sentences than female judges did (*bs* between 11.3 and 28.5). Most variables measuring the judges' perceptions and beliefs did not appear to be statistically significant predictors of the sentence. This was mostly due to the standard errors, as some of the coefficients were visibly larger than others. Overall, judges who believed the offender would bear the main responsibility for crimes meted out virtually the same sentences as those who believed that the offender and the society would each bear half of the responsibility. For the offenders versus law-abiding people question, the sentence gaps between judges who selected "very different" and "somewhat different" (*bs* between -32.3 and 17.1) were larger than those between judges who selected "somewhat different" and "basically the same" (*bs* between -8.8 and 11.2). For the two other comparison questions, the sentence gaps between judges who selected "somewhat different" and "basically the same" were larger than those between judges who selected "very different" and "somewhat different." Judges with a firmer belief in prisons' punitive effectiveness tended to mete out *longer* sentences (*bs* between -10.1 and 3.1), whereas judges with a firmer belief in prisons' rehabilitative effectiveness

tended to mete out *shorter* sentences (*bs* between -8.7 and 12.7), although these relationships were non-significant as well due to the large standard errors.<sup>8</sup>

### ***Comparison of the Perception and Belief Variables' Effects***

We next ran sex-specific models (i.e., limiting the subsample to one sex at a time) to examine if the effect of the perception and belief variables varied between sexes, and present the results in Table 3. Every two columns in Table 3 present a pair—the coefficients for the perception/belief variables for a male defendant and the corresponding female defendant. Similar to what we found in the main models, most of the coefficients were non-significant in their respective models, and had wide standard errors. Although many of the coefficient pairs were visibly different between the two vignette versions, a series of more formal interaction effect tests failed to find significant differences in these coefficients between sexes.<sup>9</sup> As a result, we are unable to conclude that the perception and belief variables had much unique effect on one defendant sex over the other.

### ***Sensitivity Checks***

**Log-transformed dependent variables.** In three out of the four vignettes (except for the homicide scenario), the distribution of the sentence length variable was skewed. We re-estimated our main models with a log-transformed sentence length variable, and present the findings in Table 4.<sup>10</sup> Similar to our main analysis,

---

<sup>8</sup> Our coding scheme for both variables here were 1 = strongly agree and 4 = strongly disagree, so each additional point on the scale was one point away from the belief in prisons' functions.

<sup>9</sup> We first added each and only one sex\*perception (responsible party, the three perceived differences, the two beliefs in prisons' functions) interaction term to our main model at a time, then also estimated one overall model with all interaction effects included, neither found noticeable interaction effects.

<sup>10</sup> For the first two sensitivity tests, we only present the coefficient for the male defendant since the key purpose was to examine if the sex disparities remained. We present the full parsimonious

we found a sex disparity favoring men for all three crime types, and the statistical significance remained the same for both robbery (non-significant) and defrauding vignettes ( $b = -0.15$ , suggesting the average sentence for the female defendant was about 14% shorter than that for the male defendant). However, the disparity for the drug vignette became non-significant. Nevertheless, our key finding—the magnitude and direction of sex disparities depending on the crime type—was still present.

**Heckman correction for the homicide vignette.** While we did not observe a major skewness issue with the sentence for the homicide vignette, over a third of judges recommended life imprisonment or the suspended death penalty, therefore having a potential selection issue. We conducted a Heckman two-step correction to address this issue, and found a post-correction coefficient for the male defendant at 71.80 (i.e., after correcting for selection bias, the male defendant received an average sentence 72 months longer than the female defendant did). Although the size of the estimated disparity was large, it was again not statistically significant due to a large standard error. Bushway et al. (2007) and Leung and Yu (1996) both cautioned against the use of Heckman correction without an exclusion restriction (i.e., the same set of independent variables predicting both selection and the ultimate outcome, which is our situation) because such practice can lead to multicollinearity issues. Although we did not see direct evidence of multicollinearity through a variance inflation factor (VIF, lower than 5.27 for all

---

models, because the purpose was to examine whether the reduction of parameters affected all coefficients.

regressors in the main model) analysis, the standard error does appear to be consistent with the main symptom of collinearity.

**Parsimonious models.** Our last sensitivity check addressed the concern that the main analysis contained too many parameters. Given our small  $n$ , each additional independent variable takes away a larger proportion of degrees of freedom than in models with a large  $n$ . We recoded both court level (1 = intermediate or higher court, 0 = primary court) and crime responsibility (1 = the offender bears the primary responsibility, 0 = otherwise) variables into dummies. We also combined the other perception and belief variables into two new variables: overall perceived difference (average score of three perceived difference variables, 1 = very different, 3 = almost the same), overall belief in prisons' effectiveness (average score of belief in punishment effectiveness and belief in rehabilitation effectiveness, 1 = strongly agree, 4 = strongly disagree).<sup>11</sup> We present the results in Table 4 as well. The sex disparity estimates—direction, magnitude, and statistical significance—were very similar to our main models. Although this approach reduced the number of parameters considerably, nearly all perception variables remained non-significant, which suggests that reducing the number of parameters alone did not appear to solve the model fitting issues.

### ***Sensitivity Power Analysis***

---

<sup>11</sup> If a judge selected “does not know/not sure” for one or more of the perceived difference questions (17 total), we dropped the judge’s response to that question, and used their response to the other perceived difference questions with a valid value to calculate the average score. For example, if a judge answered “very different” (1) to one question, “almost the same” (3) to another question, and “not sure” (4) to a third question, the average score would be 2 rather than 2.67. We also estimated the models using only judges who gave a response other than “not sure” for all questions and did not find meaningful differences in the results.

One possible reason why we saw very few statistically significant coefficients was insufficient statistical power. The concern is valid given our small  $n$  and our small-to-moderate effect sizes. To investigate the extent of the power issue, we conducted a sensitivity power analysis for regression coefficients (calculation of critical effect size with given power, alpha-level,  $n$ , and the number of regressors), using the software G\*Power 3.1 (Faul et al., 2009). For two-tailed tests at a power of .8, alpha-level of .05, and our main model's  $n$  (243) and number of parameters (17), the critical  $f^2$  value is 0.0326. Our main models'  $R^2$  values ranged between .17 (homicide) and .08 (defrauding), which converts to critical additional  $R^2$  values at around .032 (Cohen, 1988, p. 410). All of our perception and belief variables only had an additional  $R^2$  value around .01, suggesting that we may not have had the sufficient effect size to achieve a desired power level.

### **Summary of Findings**

Our findings supported the hypotheses on sex disparities. In three of our four vignettes, we found a significant disparity in the judges' recommended sentences. Judges recommended a harsher average sentence for the male homicide defendant, and recommended a harsher average sentence for the female defrauding and drug trafficking defendants. There was no significant sex disparity in the sentencing of the robbery defendant. However, there was barely any support of our hypotheses on the belief variables—either their own effect or their role as a confounder behind the sex disparities. Sensitivity checks suggest that the findings above were generally stable across different model specifications. A power analysis found that the effect sizes of the perception and

belief variables were lower than what is needed to achieve a satisfactory power level.

### Discussion

Following his critique of the “modal approach” of sentencing research in the United States, Baumer (2013) called for the next generation of research as follows,

What types of research along these lines might be pursued? One possibility... was to conduct field experiments of various sorts.... Closer to home, Paternoster and Brame suggest [that]... “it is not inconceivable to conduct an actual experiment in which prosecutors would be asked to decide whether they would seek a death sentence after reading a hypothetical case record of the homicide with the perceived race of the defendant and victim experimentally manipulated.” (p. 252, all citations omitted)

Although the focus of the present study is neither racial disparities nor the United States, we heed the call nevertheless. We brought together the feminist theories and the theories of case processing in the Chinese context, featuring a unique and innovative research design. To the best of our knowledge, the recent literature has featured no quantitative research on sentencing in China that did not rely on official data or the coding of official case verdicts. Inspired by a recent wave of experiments in the court context (Aspinwall et al., 2012; Bushway et al., 2014; Wilford et al., in press), we manipulated the sex of defendants across four vignettes, and found strong main effects and promises for future studies down this line. Of course, we have no intention to nullify the progress made by studies using

official case data—given the overall scarcity of attention to Chinese courts, every reasonable endeavor deserves its credit. That said, we hope our study reveals the possibility for future scholars to blaze a different trail.

The main finding of our study was that both the presence and the direction of sex disparity highly depended on the crime type. There was neither full support of the original chivalry hypothesis nor full support of the “evil woman” hypothesis. Instead, the results gave us a mixed message leaning towards a selective chivalry story (Farnworth & Teske, 1995; Spohn & Beichner, 2000). In the homicide vignette, judges recommended harsher sentences for the male defendant. In the other three vignettes, judges recommended harsher sentences for the female defendant, with two of the three disparities (for defrauding and drug trafficking vignettes) being relatively large and statistically significant. We depicted the homicide vignette under a domestic violence context—the offender received prolonged abuse from the victim. It is not surprising that judges, in general, considered the male defendant who killed his wife to be more blameworthy than the female defendant who killed her husband. Arguably, few scenarios could have highlighted the “less culpable, less likely to recidivate, and more amenable to rehabilitation” stereotype (Spohn & Beichner, 2000, p. 175) associated with female defendants better than a battered wife (also Daly & Tonry, 1997; Gaub & Holtfreter, 2015). In both American and Chinese contexts, it is not rare for criminal justice system actors to view women who injured or even killed their abusive partner as half an offender and half a victim (Miller, 2001; Xing, 2013; Zhao et al., 2018). Our findings echo both theories and empirical findings on similarly-situated cases (Bond & Jeffries, 2014; Lu et al., 2013; Niu & Laidler, 2015; Xing, 2013).

As we mentioned earlier in the paper, domestic violence is probably one of the crime types with the clearest gender norms and expectations. For other crime types, the norms can be fluid, and we did see different patterns empirically. For the robbery vignette, the sentence disparity between sexes was non-significant and small in absolute terms. One possible explanation is that the serious nature of robbery took away the judges' room for discretion. Previous studies found that extralegal disparities were more likely to appear in low-seriousness crimes (Spohn & Cederblom, 1991). Not only is robbery serious, robbery committed by women in reality is so rare that empirical studies found little stereotyping associated with the sex of the defendant (McKimmie et al., 2013; Spohn & Spears, 1997). We found harsher average recommended sentences for women in the other two vignettes, which contradicts the findings of studies on the sentencing of fraud (Gaub & Holtfreter, 2015; Holtfreter, 2013) and drug defendants (Liang et al., 2009; Lu & Liang, 2008). The gender norms on these issues are also less clear than those in the family context—there has been little scholarly discussion on whether women are more likely to get away from these crimes. The best evidence we have is peripheral—it is possible that the Chinese society still views women as overall more approachable, trustworthy, and cautious of the health of others, and women who defraud others or sell drugs are committing a more reprehensible violation of the gender norms relative to men who engage in the same acts. Yet before we can reach any conclusion in an academically rigorous manner, it is necessary for researchers to expand the study of gender norms and expectations onto a wider range of behaviors.

Although theories have highlighted the role of judicial beliefs and perceptions (Carroll et al., 1987; Redding & Hensl, 2011), we did not find the variables we used playing a major role in explaining the disparities found from our vignettes. The sensitivity power analysis we conducted revealed that our measures of perceptions and beliefs did not achieve the necessary effect size threshold, suggesting that these variables did not matter as much as we originally expected. Previous studies on Chinese courts have highlighted the unique form of courtroom culture and norms. In explaining their null findings on the sex of judges in China, Wei and Xiong (2020) attributed the uniformity of sentence decisions to the formal (sentencing guidelines and adjudication committees) and informal (the courtroom workgroup consisting of the police, prosecutors, and judges, known as “the iron triangle” in China) check and balance mechanisms (see also Ahl, 2014; Roberts & Pei, 2016). The same could be happening here—although judges had their own beliefs on the cause of crimes and the efficacy of the court system (and felt comfortable expressing them in the survey), the decision of the sentences relied on a different (and probably *orthogonal*) set of variables. A theme in the Chinese court literature is that individual judges’ discretion and preference can be at odds with the professional training and the structural restrictions (Liebman, 2015; Zheng et al., 2017). When local courtroom norms require handling of cases in a non-sex-neutral manner (to cope with matters such as the increase in women’s involvement in drug crimes), we can expect two outcomes to happen at the same time: sex disparities in the case outcomes, and judges’ beliefs as irrelevant.

The primary implication of the study is that extralegal disparities exist, and given the sheer amount of differences among the judges' responses to nearly all questions, the decision-making process of judges is anything but uniform. In the history of both the United States and China, there has been wide disagreements on the optimal amount of discretion at individual judges' hands, and laws and policies often swing between the ends (Ahl, 2014; Baumer, 2013). The sentence disparity for two vignettes were close to two years (drug trafficking) and four years (homicide) respectively, which deserves attention. Although our findings per se do not answer the question of how much judge-level discretion *should* be there, our findings can serve as a conversation starter for policymakers. The study also demonstrates the value of judges as research participants, and their involvement is crucial in evidence-based policies.

Overall, our study served as a bridge between apparently distant realms of empirical research: the legal and social contexts between the United States and China, and the intellectual frameworks between the criminological (sex disparities) and social psychological (perceptions and beliefs) literature. Some of our findings are consistent with the literature, while others are unique: we found evidence for selective chivalry or the dependence of leniency towards women on the crime type. However, the specific types of cases that received chivalry were different from the literature. To our awareness, the present study is among the first attempts to study Chinese judges' behaviors under the experimental framework. We hope the study paved the way for future research on courtroom decision-making in China.

In addition to the issues already discussed in the paper (such as the response rate or the effect size and statistical power), our studies have some other limitations. First, we did not manipulate the nature and specifics of the crimes. Therefore, it is difficult to decipher whether the sex disparities were due to the judges' perception of overall gender roles and risk, or specific aspects of the crimes. For example, we do not know whether the sex disparity favoring the female homicide defendant will stand if she killed a stranger as a hired hitter rather than an abusive husband, or whether the female drug defendant will face a more favorable treatment than the male if both engaged in drug trafficking to cover the medical cost of an ill family member. Second, we did not include a benchmark measure of the judges' overall punitiveness. For example, some judges may prefer to mete out harsh sentences for all violent crime defendants and less harsh sentences for all property and drug crime defendants, regardless of their sex. If these judges were overrepresented in our male defendants vignette group, then our main findings would have been an artifact due to non-random judge assignment. Some overall questions measuring the judges' punitive attitude towards different crime types would help addressing the issue. Third, although we attempted several ways to analyze the sentence for the homicide defendant, we were unable to further take advantage of the recent advances in penal metrics theory and model properties such as the diminishing return of punishment (i.e., the first year in prison being harsher than the seventh year, cf. Pina-Sánchez & Gosling, 2020; Yan, 2020). Most studies in this area focused on the conversion among sanctions on the lenient end (e.g., between prison and probation) rather than the harsh end (e.g., between fixed-term imprisonment and life imprisonment),

the latter of which can be further explored by future research. Fourth, we used a clear-cut man-woman binary for our sex variables, which was not the most inclusive measurement. Although the binary was common in sentencing research, extant literature suggests that gender identity can develop along one or multiple dimensions. Future studies will also benefit from including additional options for the gender variable.

At the conclusion of their pioneering study on sex disparity in sentencing in China, Lu et al. (2013, p. 175) left a hopeful remark that “[g]iven the professionalization and legalization movement in the legal system... and also the increasingly equal roles played by women in the Chinese society, the reactions of the criminal justice system to female offenders and their offenses will continue to evolve.” Despite the limitations, we hope that our study can serve as a proof of concept of this evolution—one that demonstrates the utility and potential of primary data collection. At its nascency, the People’s Republic brought up the idea that “women hold up half of the sky.” Both its history and the ongoing, worldwide calls for women’s rights make the issue of sex disparity in courts too big to neglect.

## References

- Ahl, B. (2014). Retaining judicial professionalism: The new guiding cases mechanism of the Supreme People's Court. *The China Quarterly*, 217, 121-139. <https://doi.org/10.1017/s0305741013001471>
- Ahl, B., & Sprick, D. (2018). Towards judicial transparency in China: The new public access database for court decisions. *China Information*, 32(1), 3-22. <https://doi.org/10.1177/0920203x17744544>
- Albonetti, C. A. (1991). An integration of theories to explain judicial discretion. *Social Problems*, 38(2), 247-266. <https://doi.org/10.2307/800532>
- Armstrong, G. (1977). Females under the law—"protected" but unequal. *Crime & Delinquency*, 23(2), 109-120. <https://doi.org/10.1177/001112877702300202>
- Aspinwall, L. G., Brown, T. R., & Tabery, J. (2012). The double-edged sword: Does biomechanism increase or decrease judges' sentencing of psychopaths? *Science*, 337(6096), 846-849. <https://doi.org/10.1126/science.1219569>
- Bai, J. (2013). 司法潜见对定罪过程的影响 [Effects of Judicial Pre-judgment on Conviction]. *中国社会科学* [Social Sciences in China], 2013(1), 167-185.
- Baumer, E. P. (2013). Reassessing and redirecting research on race and sentencing. *Justice Quarterly*, 30(2), 231-261. <https://doi.org/10.1080/07418825.2012.682602>
- Bond, C. E. W., & Jeffries, S. (2014). Similar punishment? Comparing sentencing outcomes in domestic and non-domestic violence cases. *The British Journal of Criminology*, 54(5), 849-872. <https://doi.org/10.1093/bjc/azu034>
- Bontrager, S., Barrick, K., & Stupi, E. (2013). Gender and sentencing: A meta-analysis of contemporary research. *Journal of Gender, Race, & Justice*, 16(2), 349-372.
- Bumby, K. M., & Maddox, M. C. (1999). Judges' knowledge about sexual offenders, difficulties presiding over sexual offense cases, and opinions on sentencing, treatment, and legislation. *Sexual Abuse: A Journal of Research and Treatment*, 11(4), 305-315. <https://doi.org/10.1023/A:1021367015037>
- Bushway, S. D., Johnson, B. D., & Slocum, L. A. (2007). Is the magic still there? The use of the Heckman two-step correction for selection bias in criminology. *Journal of Quantitative Criminology*, 23(2), 151-178. <https://doi.org/10.1007/s10940-007-9024-4>

- Bushway, S. D., Redlich, A. D., & Norris, R. J. (2014). An explicit test of plea bargaining in the "shadow of the trial". *Criminology*, 52(4), 723-754. <https://doi.org/10.1111/1745-9125.12054>
- Bustelo, M. (2009). Spain: Intersectionality faces the strong gender norm. *International Feminist Journal of Politics*, 11(4), 530-546. <https://doi.org/10.1080/14616740903237491>
- Carroll, J. S., Perrowitz, W. T., Lurigio, A. J., & Weaver, F. M. (1987). Sentencing goals, causal attributions, ideology, and personality. *Journal of Personality and Social Psychology*, 52(1), 107-118. <https://doi.org/10.1037/0022-3514.52.1.107>
- Cassidy, M., & Rydberg, J. (2020). Does sentence type and length matter? Interactions of age, race, ethnicity, and gender on jail and prison sentences. *Criminal Justice and Behavior*, 47(1), 61-79. <https://doi.org/10.1177/0093854819874090>
- Chesney-Lind, M. (2006). Patriarchy, crime, and justice: Feminist criminology in an era of backlash. *Feminist Criminology*, 1(1), 6-26. <https://doi.org/10.1177/1557085105282893>
- Chesney-Lind, M. (2020). Feminist criminology in an era of misogyny. *Criminology*, Advance Online Publication. <https://doi.org/10.1111/1745-9125.12247>
- Chesney-Lind, M., & Pasko, L. (2013). *The female offender: Girls, women, and crime* (3rd ed.). SAGE.
- Christensen, L. S. (2018). The New Portrayal of Female Child Sexual Offenders in the Print Media: A Qualitative Content Analysis. *Sexuality & Culture*, 22(1), 176-189. <https://doi.org/10.1007/s12119-017-9459-1>
- Clarke, D. (2003). Empirical research into the Chinese judicial system. In E. G. Jensen & T. C. Heller (Eds.), *Beyond common knowledge: Empirical approaches to the rule of law* (pp. 164-192). Stanford University Press.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Lawrence Erlbaum Associates.
- Croll, E. (1978). *Feminism and socialism in China*. Routledge & Kegan Paul Ltd.
- Daly, K., & Bordt, R. L. (1995). Sex effects and sentencing: An analysis of the statistical literature. *Justice Quarterly*, 12(1), 141-175. <https://doi.org/10.1080/07418829500092601>
- Daly, K., & Chesney-Lind, M. (1988). Feminism and criminology. *Justice Quarterly*, 5(4), 497-538. <https://doi.org/10.1080/07418828800089871>

- Daly, K., & Tonry, M. (1997). Gender, race, and sentencing. *Crime and Justice*, 22, 201-252. <https://doi.org/10.1086/449263>
- Drápal, J. (2020). Sentencing disparities in the Czech Republic: Empirical evidence from post-communist Europe. *European Journal of Criminology*, 17(2), 151-174. <https://doi.org/10.1177/1477370818773612>
- Embry, R., & Lyons, P. M. (2012). Sex-based sentencing: Sentencing discrepancies between male and female sex offenders. *Feminist Criminology*, 7(2), 146-162. <https://doi.org/10.1177/1557085111430214>
- Etienne, M. (2010). Sentencing women: Reassessing the claims of disparity. *Journal of Gender, Race, & Justice*, 14(1), 73-84.
- Farnworth, M., & Teske, R. H. C., Jr. (1995). Gender differences in felony court processing. *Women & Criminal Justice*, 6(2), 23-44. [https://doi.org/10.1300/J012v06n02\\_02](https://doi.org/10.1300/J012v06n02_02)
- Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using G\*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods*, 41(4), 1149-1160. <https://doi.org/10.3758/BRM.41.4.1149>
- Finckenauer, J. O. (1978). Crime as a national political issue: 1964-76. *Crime & Delinquency*, 24(1), 13-27. <https://doi.org/10.1177/001112877802400102>
- Fishbane, A., Ouss, A., & Shah, A. K. (2020). Behavioral nudges reduce failure to appear for court. *Science*, 370(6517). <https://doi.org/10.1126/science.abb6591>
- Frase, R. S. (1997). Sentencing principles in theory and practice. *Crime and Justice*, 22, 363-433. <https://doi.org/10.1086/449266>
- Fridel, E. E. (2019). Leniency for lethal ladies: Using the actor-partner interdependence model to examine gender-based sentencing disparities. *Homicide Studies*, 23(4), 319-343. <https://doi.org/10.1177/1088767919867420>
- Gaub, J. E., & Holtfreter, K. (2015). New directions in intersections, inequality, and sentencing. *Women & Criminal Justice*, 25(5), 298-312. <https://doi.org/10.1080/08974454.2014.989299>
- Gil, V. E., & Anderson, A. F. (1999). Case study of rape in contemporary China. *Journal of Interpersonal Violence*, 14(11), 1151-1171. <https://doi.org/10.1177/088626099014011003>

- Goulette, N. (2021). Is chivalry dead? An examination of charge reductions in one Ohio jurisdiction. *Women & Criminal Justice*, 31(2), 142-158.  
<https://doi.org/10.1080/08974454.2020.1753141>
- Grabe, M. E., Trager, K. D., Lear, M., & Rauch, J. (2006). Gender in crime news: A case study test of the chivalry hypothesis. *Mass Communication and Society*, 9(2), 137-163. [https://doi.org/10.1207/s15327825mcs0902\\_2](https://doi.org/10.1207/s15327825mcs0902_2)
- Hayes, A. F., Shanahan, J., & Glynn, C. J. (2001). Willingness to express one's opinion in a realistic situation as a function of perceived support for that opinion. *International Journal of Public Opinion Research*, 13(1), 45-58.  
<https://doi.org/10.1093/ijpor/13.1.45>
- Hill, R. A. (1998). Character, choice, and "aberrant behavior:" Aligning criminal sentencing with concepts of moral blame. *The University of Chicago Law Review*, 65(3), 975-999. <https://doi.org/10.2307/1600304>
- Ho, S. S., & McLeod, D. M. (2008). Social-Psychological Influences on Opinion Expression in Face-to-Face and Computer-Mediated Communication. *Communication Research*, 35(2), 190-207.  
<https://doi.org/10.1177/0093650207313159>
- Holtfreter, K. (2013). Gender and "other people's money": An analysis of white-collar offender sentencing. *Women & Criminal Justice*, 23(4), 326-344.  
<https://doi.org/10.1080/08974454.2013.821015>
- Kaufman, J. P., & Williams, K. P. (2013). *Women at war, women building peace: Challenging gender norms*. Kumarian Press.
- Kinkel, J. J., & Hurst, W. J. (2015). The judicial cadre evaluation system in China: From quantification to intra-state legibility. *The China Quarterly*, 224, 933-954. <https://doi.org/10.1017/s0305741015001290>
- Leung, S. F., & Yu, S. (1996). On the choice between sample selection and two-part models. *Journal of Econometrics*, 72(1-2), 197-229.  
[https://doi.org/10.1016/0304-4076\(94\)01720-4](https://doi.org/10.1016/0304-4076(94)01720-4)
- Liang, B., Lu, H., & Taylor, M. (2009). Female drug abusers, narcotics offenders, and legal punishment in China. *Journal of Criminal Justice*, 37(2), 133-141.  
<https://doi.org/10.1016/j.jcrimjus.2009.02.007>
- Liebman, B. L. (2015). Leniency in chinese criminal law? Everyday justice in Henan. *Berkeley Journal of International Law*, 33(1), 153-222.  
<https://doi.org/10.15779/Z38NP20>
- Liu, M. (2012). Chinese migrant women in the sex industry. *Feminist Criminology*, 7(4), 327-349. <https://doi.org/10.1177/1557085112436836>

- Liu, S., Xiong, M., & Liang, B. (2018). The death penalty for foreign drug offenders in China: Legal protection and equal treatment. *European Journal on Criminal Policy and Research*, 25(4), 427-448. <https://doi.org/10.1007/s10610-018-9377-2>
- Lu, H., & Liang, B. (2008). Legal responses to trafficking in narcotics and other narcotic offenses in China. *International Criminal Justice Review*, 18(2), 212-228. <https://doi.org/10.1177/1057567708318480>
- Lu, H., Liang, B., & Liu, S. (2013). Serious violent offenses and sentencing decisions in China-Are there any gender disparities? *Asian Journal of Criminology*, 8(3), 159-177. <https://doi.org/10.1007/s11417-012-9155-x>
- MacKinnon, C. A. (1983). Feminism, Marxism, method, and the state: Toward feminist jurisprudence. *Signs: Journal of Women in Culture and Society*, 8(4), 635-658. <https://doi.org/10.1086/494000>
- Mann, S. L. (2011). *Gender and sexuality in modern Chinese history*. Cambridge University Press.
- Martin, N. D., Rigoni, D., & Vohs, K. D. (2017). Free will beliefs predict attitudes toward unethical behavior and criminal punishment. *Proceedings of the National Academy of Sciences of the United States of America*, 114(28), 7325-7330. <https://doi.org/10.1073/pnas.1702119114>
- McKimmie, B. M., Masters, J. M., Masser, B. M., Schuller, R. A., & Terry, D. J. (2013). Stereotypical and counterstereotypical defendants: Who is he and what was the case against her? *Psychology, Public Policy, and Law*, 19(3), 343-354. <https://doi.org/10.1037/a0030505>
- Miller, S. L. (2001). The paradox of women arrested for domestic violence. *Violence Against Women*, 7(12), 1339-1376. <https://doi.org/10.1177/10778010122183900>
- Niu, X., & Laidler, K. A. J. (2015). Understanding domestic violence against muslim women in China. *Feminist Criminology*, 10(1), 92-112. <https://doi.org/10.1177/1557085114536766>
- Nowacki, J. S. (2019). Gender equality and sentencing outcomes: An examination of state courts. *Criminal Justice Policy Review*, Advance online publication. <https://doi.org/10.1177/0887403419840804>
- Pina-Sánchez, J., & Gosling, J. P. (2020). Tackling selection bias in sentencing data analysis: a new approach based on a scale of severity. *Quality & Quantity*, 54(3), 1047-1073. <https://doi.org/10.1007/s11135-020-00973-z>

- Pina-Sánchez, J., & Harris, L. (2020). Sentencing gender? Investigating the presence of gender disparities in Crown Court sentences. *Criminal Law Review*, 2020(1), 3-28.
- Pittaway, E., & Bartolomei, L. (2001). Refugees, race, and gender: The multiple discrimination against refugee women. *Refuge: Canada's Journal on Refugees*, 19(6), 21-32.
- Rapaport, E. (1991). The death penalty and gender discrimination. *Law & Society Review*, 25(2), 367-383. <https://doi.org/10.2307/3053803>
- Redding, R. E., & Hensl, K. B. (2011). Knowledgeable judges make a difference: Judicial beliefs affect juvenile court transfer decisions. *Juvenile and Family Court Journal*, 62(3), 15-24. <https://doi.org/10.1111/j.1755-6988.2011.01063.x>
- Roberts, J. V., & Pei, W. (2016). Structuring judicial discretion in china: Exploring the 2014 sentencing guidelines. *Criminal Law Forum*, 27(1), 3-33. <https://doi.org/10.1007/s10609-015-9270-3>
- Rodriguez, S. F., Curry, T. R., & Lee, G. (2006). Gender differences in criminal sentencing: Do effects vary across violent, property, and drug offenses? *Social Science Quarterly*, 87(2), 318-339. <https://doi.org/10.1111/j.1540-6237.2006.00383.x>
- Ruskola, T. (1994). Law, sexual morality, and gender equality in Qing and communist China. *The Yale Law Journal*, 103(8), 2531-2565. <https://doi.org/10.2307/797055>
- Santos, G., & Harrell, S. (Eds.). (2017). *Transforming patriarchy: Chinese families in the twenty-first century*. University of Washington Press.
- Shariff, A. F., Greene, J. D., Karremans, J. C., Luguri, J. B., Clark, C. J., Schooler, J. W., Baumeister, R. F., & Vohs, K. D. (2014). Free will and punishment: A mechanistic view of human nature reduces retribution. *Psychological Science*, 25(8), 1563-1570. <https://doi.org/10.1177/0956797614534693>
- Shen, A., & Winlow, S. (2014). Women and crime in contemporary China: A review essay. *International Journal of Comparative and Applied Criminal Justice*, 38(4), 327-342. <https://doi.org/10.1080/01924036.2013.861354>
- Smith, T., & Jiang, S. (2019). Making sense of life without parole in China. *Punishment & Society*, 21(1), 70-88. <https://doi.org/10.1177/1462474517739848>

- Spamann, H., & Klöhn, L. (2016). Justice is less blind, and less legalistic, than we thought: Evidence from an experiment with real judges. *Journal of Legal Studies*, 45(2), 255-280. <https://doi.org/10.1086/688861>
- Spohn, C. (1999). Gender and sentencing of drug offenders: Is chivalry dead? *Criminal Justice Policy Review*, 9(3-4), 365-399. <https://doi.org/10.1177/088740349900900305>
- Spohn, C., & Beichner, D. (2000). Is preferential treatment of female offenders a thing of the past? A multisite study of gender, race, and imprisonment. *Criminal Justice Policy Review*, 11(2), 149-184. <https://doi.org/10.1177/0887403400011002004>
- Spohn, C., & Cederblom, J. (1991). Race and disparities in sentencing: A test of the liberation hypothesis. *Justice Quarterly*, 8(3), 305-327. <https://doi.org/10.1080/07418829100091071>
- Spohn, C., Welch, S., & Gruhl, J. (1985). Women defendants in court: The interaction between sex and race in convicting and sentencing. *Social Science Quarterly*, 66(1), 178-185.
- Spohn, C. C., & Spears, J. W. (1997). Gender and case processing decisions. *Women & Criminal Justice*, 8(3), 29-59. [https://doi.org/10.1300/J012v08n03\\_02](https://doi.org/10.1300/J012v08n03_02)
- Steffensmeier, D., Ulmer, J. T., & Kramer, J. H. (1998). The interaction of race, gender, and age in criminal sentencing: The punishment cost of being young, black, and male. *Criminology*, 36(4), 763-798. <https://doi.org/10.1111/j.1745-9125.1998.tb01265.x>
- Steffensmeier, D. J. (1980). Assessing the impact of the women's movement on sex-based differences in the handling of adult criminal defendants. *Crime & Delinquency*, 26(3), 344-357. <https://doi.org/10.1177/001112878002600305>
- Tillyer, R., Hartley, R. D., & Ward, J. T. (2015). Differential treatment of female defendants: Does criminal history moderate the effect of gender on sentence length in federal narcotics cases? *Criminal Justice and Behavior*, 42(7), 703-721. <https://doi.org/10.1177/0093854814560624>
- Trevaskes, S. (2013). China's death penalty: The Supreme People's Court, the suspended death sentence and the politics of penal reform. *The British Journal of Criminology*, 53(3), 482-499. <https://doi.org/10.1093/bjc/azt002>

- Ulmer, J. T. (2012). Recent developments and new directions in sentencing research. *Justice Quarterly*, 29(1), 1-40.  
<https://doi.org/10.1080/07418825.2011.624115>
- Wei, S., & Xiong, M. (2020). Judges' gender and sentencing in China: An empirical inquiry. *Feminist Criminology*, 15(2), 217-250.  
<https://doi.org/10.1177/1557085119842660>
- Wilford, M. M., Zimmerman, D. M., Yan, S., & Sutherland, K. T. (in press). Innocence in the shadow of COVID-19: Plea decision making during a pandemic. *Journal of Experimental Psychology: Applied*.  
<https://doi.org/10.1037/xap0000367>
- Wu, Y. (2020a). Is a plea really a bargain? An empirical study of six cities in China. *Asian Journal of Criminology*, 15(3), 237-258.  
<https://doi.org/10.1007/s11417-020-09322-3>
- Wu, Y. (2020b). On the effect of the Chinese version of speedy trial and Plea bargaining pilot programs: Observation from DUI cases in Fujian Province. *Crime, Law and Social Change*, 74(4), 457-484.  
<https://doi.org/10.1007/s10611-020-09904-3>
- Xia, Y., Cai, T., & Zhong, H. (2019). Effect of judges' gender on rape sentencing: A data mining approach to analyze judgment documents. *The China Review*, 19(2), 125-149.
- Xing, H. (2013). 论配偶暴力中受虐妇女杀夫案的量刑 [Rethinking punishment for women who are driven by domestic violence to kill]. *中华女子学院学报* [*Journal of China Women's University*], 2013(1), 24-28.  
<https://doi.org/10.3969/j.issn.1007-3698.2013.01.004>
- Xiong, M., Liu, S., & Liang, B. (2018). Criminal defense and judicial sentencing in China's death penalty cases. *Psychology, Crime & Law*, 24(4), 414-432.  
<https://doi.org/10.1080/1068316x.2017.1390114>
- Yan, S. (2020). What exactly is the bargain? The sensitivity of plea discount estimates. *Justice Quarterly*, Advance online publication.  
<https://doi.org/10.1080/07418825.2019.1707856>
- Yan, Y. (2006). Girl power: Young women and the waning of patriarchy in rural north China. *Ethnology*, 45(2). <https://doi.org/10.2307/4617569>
- Zatz, M. S. (2000). The convergence of race, ethnicity, gender, and class on court decisionmaking: Looking toward the 21st century. In J. Horney (Ed.), *Criminal justice 2000: Policies, processes, and decisions of the criminal justice system* (Vol. 3). National Institute of Justice.

- Zhao, R., Zhang, H., Jiang, Y., & Yao, X. (2018). The tendency to make arrests in domestic violence: Perceptions from police officers in China. *International Journal of Offender Therapy and Comparative Criminology*, 62(16), 4923-4941. <https://doi.org/10.1177/0306624X18801653>
- Zheng, C., Ai, J., & Liu, S. (2017). The elastic ceiling: Gender and professional career in Chinese courts. *Law & Society Review*, 51(1), 168-199. <https://doi.org/10.1111/lasr.12249>

Table 1. Descriptive Statistics

	<i>n</i>	<i>M</i>	<i>SD</i>	Min	Max
Judge characteristics					
Male	272	0.68	0.47		
Graduate degree	272	0.34	0.48		
Age	260	39.97	10.41	20	60
Primary court	271	0.39	0.49		
Intermediate court	271	0.37	0.48		
Higher court	271	0.23	0.42		
Perceived responsible party of crimes					
Offenders primarily responsible	269	0.42	0.49		
Offenders and society equally responsible	269	0.57	0.50		
Society primarily responsible	269	0.01	0.11		
Perception of offenders and law-abiding people					
Very different	269	0.09	0.29		
Somewhat different	269	0.50	0.50		
Not different	269	0.30	0.46		
Does not know	269	0.11	0.32		
Perception of offenders committing different types of crime					
Very different	267	0.28	0.45		
Somewhat different	267	0.51	0.50		
Not different	267	0.13	0.34		
Does not know	267	0.08	0.27		
Perception of male and female offenders					
Very different	271	0.33	0.47		
Somewhat different	271	0.45	0.50		
Not different	271	0.15	0.36		
Does not know	271	0.07	0.25		
Belief that prisons can punish	269	2.32	0.58	1	4
Belief that prisons can rehabilitate	271	2.65	0.676	1	4
Male defendant vignettes	272	0.43	0.50		
Sentence for homicide					
Fixed-term imprisonment	271	0.65	0.48		
Life imprisonment	271	0.28	0.45		

Death with suspension	271	0.07	0.26		
Length for fixed-term imprisonment	170	114.08	45.77	24	180
Sentence for robbery					
Fixed-term imprisonment	268	0.97	0.18		
Life imprisonment	268	0.03	0.18		
Length for fixed-term imprisonment	251	83.67	38.19	18	180
Sentence for defrauding					
Fixed-term imprisonment	261	0.99	0.09		
Life imprisonment	261	0.01	0.09		
Length for fixed-term imprisonment	248	44.45	21.87	6	126
Sentence for drug trafficking					
Fixed-term imprisonment	258	0.96	0.19		
Life imprisonment	258	0.02	0.14		
Death with suspension	258	0.01	0.11		
Death without suspension	258	0.01	0.09		
Length for fixed-term imprisonment	240	81.65	31.29	12	180

---

Note: Minimum and maximum values for dummy variables omitted. Statistics for all sentence length variables are for judges who recommended fixed-term imprisonment only, before conversion (for judges who recommended life imprisonment or death) and imputation (for judges who did not give out any recommended sentence or recommended fixed-term imprisonment but did not give out a length).

Table 2. Main Regression Model Explaining Sentence Length

	(1)	(2)	(3)	(4)
	Homicide	Robbery	Defrauding	Drug trafficking
Male defendant vignettes	43.80*** (12.24)	-4.90 (8.92)	-10.76* (5.25)	-20.93* (9.59)
Judge characteristics				
Male	28.53* (12.96)	22.36* (9.45)	11.28* (5.56)	11.64 (10.16)
Graduate degree	-2.02 (12.23)	13.61 (8.91)	-5.48 (5.24)	22.79* (9.58)
Age	-0.80 (0.67)	0.58 (0.49)	-0.34 (0.29)	-0.26 (0.52)
Primary court	-10.32 (13.12)	21.93* (9.57)	5.27 (5.63)	-2.72 (10.28)
Intermediate court	17.95 (15.51)	13.97 (11.31)	0.18 (6.65)	-10.73 (12.15)
Perceived responsible party of crimes (ref: offender and society equally responsible)				
Offender primarily responsible	3.64 (11.34)	3.74 (8.27)	-0.95 (4.86)	-3.17 (8.89)
Society primarily responsible	25.09 (62.26)	-63.86 (45.39)	1.10 (26.69)	-40.70 (48.79)
Perception of offenders and law-abiding people (ref: somewhat different)				
Very different	-32.32 (20.19)	-16.84 (14.72)	-14.30 (8.66)	17.11 (15.82)
Not different	-3.06 (13.78)	-4.19 (10.04)	-8.80 (5.91)	11.18 (10.80)
Not sure/doesn't know	-9.59 (19.86)	5.30 (14.48)	-4.10 (8.51)	16.28 (15.56)
Perception of offenders committing different types of crimes (ref: somewhat different)				
Very different	3.27	2.95	-2.49	7.81

	(14.22)	(10.36)	(6.09)	(11.14)
Not different	25.23	8.52	-3.17	-14.58
	(17.57)	(12.81)	(7.53)	(13.77)
Not sure/doesn't know	3.94	2.91	-0.55	6.00
	(22.10)	(16.11)	(9.47)	(17.32)
Perception of male and female offenders (ref: somewhat different)				
Very different	0.70	3.73	3.47	-3.49
	(12.83)	(9.36)	(5.50)	(10.06)
Not different	-14.21	-8.02	0.43	-0.40
	(17.47)	(12.73)	(7.49)	(13.69)
Not sure/doesn't know	-44.62	-2.36	-4.00	-1.07
	(24.59)	(17.93)	(10.54)	(19.27)
Belief that prisons can punish	-10.14	-8.00	3.14	-7.09
	(11.18)	(8.15)	(4.79)	(8.76)
Belief that prisons can rehabilitate	-8.74	10.98	2.10	12.70
	(9.64)	(7.03)	(4.13)	(7.56)
Constant	213.73***	27.40	48.88**	77.19*
	(39.43)	(28.75)	(16.90)	(30.90)
<i>n</i>	243	243	243	243
<i>R</i> <sup>2</sup>	0.17	0.11	0.08	0.11

Note: \*\*\*  $p < 0.001$ , \*\*  $p < 0.01$ , \*  $p < 0.05$

Table 3. Comparison of Sex-specific Coefficients for Perception and Belief Variables

	Homicide		Robbery		Defrauding		Drug Trafficking	
	Female	Male	Female	Male	Female	Male	Female	Male
Perceived responsible party of crimes (ref: equally responsible/ society primarily responsible)								
Offender primarily responsible	-4.31 (15.80)	14.95 (16.70)	-3.50 (10.89)	6.79 (13.11)	-2.28 (8.44)	0.75 (4.32)	-5.02 (15.97)	1.87 (5.66)
Perception of offenders and law-abiding people (ref: somewhat different)								
Very different	-37.57 (27.49)	-15.35 (31.42)	-4.54 (18.95)	-30.50 (24.66)	-12.96 (14.67)	-11.22 (8.14)	19.84 (27.78)	4.86 (10.65)
Not different	10.93 (17.97)	-9.09 (22.36)	-2.77 (12.39)	-7.92 (17.55)	-10.55 (9.59)	-2.99 (5.79)	20.82 (18.16)	-0.57 (7.58)

Not sure/doesn't know	-24.14 (29.00)	-14.15 (31.11)	0.50 (19.99)	27.42 (24.41)	-11.16 (15.48)	-0.79 (8.06)	25.58 (29.30)	-3.35 (10.54)
Perception of offenders committing different types of crimes (ref: somewhat different)								
Very different	31.33 (19.99)	-38.27 (21.29)	7.94 (13.78)	-9.70 (16.71)	-6.69 (10.67)	2.89 (5.51)	13.42 (20.20)	2.22 (7.21)
Not different	53.33* (25.44)	-9.00 (25.16)	19.78 (17.54)	4.38 (19.75)	-11.55 (13.58)	2.74 (6.52)	-11.36 (25.71)	-13.39 (8.53)
Not sure/doesn't know	8.21 (27.09)	-14.05 (41.41)	12.93 (18.67)	-22.26 (32.50)	-4.10 (14.46)	4.72 (10.72)	13.23 (27.37)	-6.72 (14.03)
Perception of male and female offenders (ref: somewhat different)								
Very different	20.17 (18.59)	-18.21 (19.60)	-0.28 (12.81)	22.07 (15.38)	9.01 (9.92)	-2.25 (5.08)	-13.23 (18.78)	-3.78 (6.64)

Not different	-31.42 (24.71)	9.34 (26.28)	-7.91 (17.03)	-3.64 (20.63)	-1.93 (13.19)	-1.25 (6.81)	-14.73 (24.97)	1.59 (8.91)
Not sure/doesn't know	-56.98* (28.32)	24.18 (56.96)	2.98 (19.52)	-10.44 (44.71)	-8.93 (15.12)	10.05 (14.75)	-8.33 (28.62)	15.35 (19.30)
Belief that prisons can punish	-6.98 (14.50)	-23.26 (18.60)	-4.48 (10.00)	-21.18 (14.60)	2.57 (7.74)	4.53 (4.82)	-11.64 (14.65)	5.58 (6.30)
Belief that prisons can rehabilitate	-6.63 (12.46)	-5.24 (16.62)	-0.15 (8.59)	36.05** (13.04)	6.67 (6.65)	-6.88 (4.30)	23.90 (12.59)	-3.48 (5.63)

Note: \*\*\*  $p < 0.001$ , \*\*  $p < 0.01$ , \*  $p < 0.05$

Table 4. Sensitivity Checks

	(1) Homicide	(2) Robbery	(3) Defrauding	(4) Drug trafficking
A. Log-transformed dependent variables				
Male defendant vignette	--	-0.03 (0.07)	-0.15* (0.07)	-0.08 (0.08)
B. Heckman correction				
Male defendant vignette	71.80 (77.80)	--	--	--
C. Parsimonious models				
Male defendant vignette	49.33*** (11.34)	-2.02 (8.56)	-10.15* (4.74)	-22.46* (8.83)
Male judge	25.50* (12.33)	16.55 (9.30)	12.41* (5.15)	12.92 (9.60)
Graduate degree	1.49 (11.62)	9.41 (8.77)	-6.15 (4.86)	21.15* (9.05)
Age of judge	-0.78 (0.62)	0.33 (0.47)	-0.44 (0.26)	-0.50 (0.48)
Intermediate or higher court	-2.72 (11.69)	20.21* (8.82)	1.75 (4.89)	-5.69 (9.10)
Offender primarily responsible	3.86 (10.88)	2.92 (8.21)	-3.79 (4.55)	-3.85 (8.47)
Average perceived difference	5.29 (10.79)	-4.56 (8.14)	-3.55 (4.51)	-7.74 (8.40)
Average belief in prisons' functions	-24.60* (10.32)	2.08 (7.78)	3.00 (4.31)	7.43 (8.03)

Note: \*\*\*  $p < 0.001$ , \*\*  $p < 0.01$ , \*  $p < 0.05$ . Panels A and B also controlled for all other variables used in main model.