

Working Paper in Economics No. 780

The Persistence of the Criminal Justice Gender Gap: Evidence from 200 Years of Judicial Decisions

Anna Bindler and Randi Hjalmarsson

Department of Economics, October 2019

ISSN 1403-2473 (Print)
ISSN 1403-2465 (Online)



UNIVERSITY OF GOTHENBURG
SCHOOL OF BUSINESS, ECONOMICS AND LAW

The Persistence of the Criminal Justice Gender Gap: Evidence from 200 Years of Judicial Decisions^{*}

Anna Bindler
University of Gothenburg

Randi Hjalmarsson
University of Gothenburg and CEPR

This version: October 16, 2019

Abstract: We document persistent gender gaps favoring females in jury convictions and judge sentences in nearly 200 years of London trials, which are unexplained by case characteristics. We find that three sharp changes in punishment severity locally affected the size and nature of the gaps, but were generally not strong enough to offset their persistence. These local effects suggest a mechanism of taste-based discrimination (paternalism) where the all-male judiciary protected females from the harshest available punishment.

JEL Codes: J16, K14, K40, N33

Keywords: gender, gender gap, crime, verdict, sentencing, discrimination, history

^{*}This paper would not have been possible without the tremendous effort of our research assistants Michael Bekele and Srinidhi Srinivasan, the generous help with the data extraction by Florin Maican, and the financial support of the Foundation for Economic Research in West Sweden, and Vetenskapsrådet, The Swedish Research Council, Grants for Distinguished Young Researchers. We thank conference and seminar participants at the University of Gothenburg, University of Amsterdam, VU Amsterdam, KIE VSE Prague, the Bank of Lithuania, the IZA workshop on gender and family economics and the University of Antwerp for many helpful comments and suggestions. Authors: Randi Hjalmarsson, Department of Economics, University of Gothenburg, Email: randi.hjalmarsson@economics.gu.se. Anna Bindler, Department of Economics, University of Gothenburg, Email: anna.bindler@economics.gu.se.

1. Introduction

In contrast to the well-known gender gaps favoring men in the labor market, a growing number of studies document a criminal justice system gender gap – especially at the sentencing stage – that treats females more leniently than males.¹ Most dramatically, Starr (2015) finds that male defendants in U.S. Federal Courts receive 63% longer sentences than females, even after conditioning on observable case characteristics. Similar gender gaps are seen in the English justice system – the subject of the current paper – today.² Besides the fundamental question of whether the courts apply equal standards across defendants of different characteristics, understanding the determinants of judicial decisions (including defendant gender) is important given the potential social and economic effects of criminal justice system interactions on both the defendant and other members of society.

The current paper contributes to the understanding of the modern-day criminal justice gender gap by studying its evolution in a dynamic period spanning the 18th and 19th centuries (1715-1900) and almost 200,000 trials at the Old Bailey Central Criminal Court of London. Our data come from *The Proceedings of the Old Bailey*, which were published after each monthly court session. This unique data source has since been digitized by *The Old Bailey Proceedings Online*, and includes (tagged) information identifying the case, session date, defendant's name and gender, detailed offense categories (more than 20 of which are non-distinctly male or female), as well as verdict and sentencing outcomes. The first part of the paper describes the gender gaps over time, while the second part demonstrates that these persistent gaps cannot be explained by observable (and proxies for unobservable) case differences. The final part of the paper, therefore, considers the role of discrimination, and exploits the dynamically changing sanction regimes to test for taste-based discrimination in the form of paternalism.

First, we document the evolution in the raw and adjusted gender gaps in jury convictions and judge sentences. Though descriptive in nature, this analysis makes two important contributions: (i) this is the first study of the dynamics of the gender gap over an extended period (let alone 200 years)³ and (ii) one of the few studies of gender gaps in multiple stages of

¹ See Blau and Kahn's (2017) review of the U.S. gender wage gap and Altonji and Blank (1999).

² Depending on the offense, men in England and Wales in 2009 were between 1.1 and 3.2 times more likely to be sentenced to immediate custody than women, and the average male to female sentence length ratio is less than one for just one of ten crime categories (criminal damage). Sentencing statistics are calculated from the data underlying a Ministry of Justice publication, Sentencing Statistics England and Wales 2009. See Tables 2i and 2j. <http://webarchive.nationalarchives.gov.uk/20140712021330/https://www.gov.uk/government/publications/sentencing-statistics-annual-ns>. Sourced from the National Archives on August 23, 2017.

³ Most existing research focuses on modern-day static snapshots in U.S. federal court data (Starr, 2015; Mustard, 2001; Schanzenbach, 2005; Sorensen et al., 2012), state courts (see Butcher et al.'s (2017) analysis of Kansas) and

the criminal justice system for the same sample of defendants.⁴ The main finding is that a significant conviction gap favoring females persists throughout the entire time period. Relative to the mean conviction rate, the gap declines until around 1850 and then increases again. Second, generalizing the sentencing patterns, we see that females are less likely to be sentenced to a particular punishment when there are more lenient sentencing options available. Conversely, females are relatively more likely to be sentenced to a particular punishment than males if that punishment was not the most severe available. Moreover, we find little (to no) evidence that the gender gap is explained by differences in observable case characteristics or even unobservable differences (as proxied by the number of words per case in the Proceedings).

What explains the historical criminal justice gender gap? The second part of the paper assesses the potential importance of taste-based discrimination by taking advantage of the dynamic nature of the criminal justice system during the 18th and 19th centuries. We use variation in punishment severity induced by three criminal justice ‘reforms’, as the society transitioned from relying on both transportation to a penal colony in the Americas and capital punishment in the mid-1700s to primarily incarceration by 1900. We treat the American Revolution in 1776 as the first natural ‘experiment’: Prior to the Revolution, about 70% of convicted defendants were sentenced to transportation to the Americas. The unexpected loss of the American penal colony resulted in a crisis in England, with a temporary (10-year) solution of incarcerating offenders in the hulks of boats on the River Thames until a new penal colony was established in Australia in 1786. Moreover, this led to the introduction of imprisonment as a sentence, which did not completely disappear upon the reinstatement of transportation. The second ‘experiment’ is the offense specific abolition of capital punishment from 1807 to 1856, and corresponding sharp rise in the prison population. Finally, penal transportation to Australia was largely abolished in 1853 (and completely in 1857). Are males and females differentially affected by these sentencing reforms, in a way which is consistent with an all-male judiciary protecting the ‘weaker’ sex, i.e. engaging in paternalism?

A consistent pattern emerges in our empirical analyses of each sentencing ‘reform’. First, we see that the pre-Revolution gender gap in convictions decreases with the evolution of prison

other countries like France (Philippe, forthcoming). There is limited research on historical criminal justice gender gaps. One exception is Bodenhorn (2009), who finds females received shorter sentences in 19th century Pennsylvania. The focus of the paper was not on gender, however, as they comprised just 4% of the 10,000 individuals sentenced. Vickers (2016) considers the effect of social status on sentencing for males in 19th century England and Wales.

⁴ An exception is Starr (2015), who studies charges, convictions and sentencing. Most previous research focused on just sentencing (Sorensen et al., 2012; Mustard, 2001; and Schanzenbach, 2005) or earlier stages, such as arrest (Stolzenberg and D’Alessio, 2004) and plea bargaining (Spohn and Spears, 1997 and Shermer and Johnson, 2010).

as a sentencing option. That is, the gender gap in convictions decreases as the *minimum* expected punishment decreases. In terms of sentencing, we find that during the temporary halt, transportation eligible offenses are shifted mainly into prison/prison hulks and corporal punishment; during the new transportation to Australia/prison regime, a significant increase in sentences to prison and corporal punishment is observed compared to the pre-Revolution period. Most importantly for testing for preference-based discrimination, we find that convicted females are disproportionately more likely (14 percentage points) to be shifted into prison during both the temporary halt and new post period. We argue that these changes represent jury and judge preferences, and not simply constraints (e.g. prison hulks excluding females), since females were also shifted out of corporal punishment, which is not capacity constrained.

Similarly, abolishing the death penalty nearly eliminated the pre-existing gender gaps in conviction. For violent offenses, the 12-percentage point gap in the chance of conviction is almost completely offset by the abolition of capital punishment. For property crimes, the gender gap in conviction of a lesser charge (about 5 percentages points) is offset by the reform. There is also some evidence that convicted females are disproportionately shifted out of transportation (as it becomes the new harshest punishment) with the abolition of capital punishment.

Finally, as transportation was abolished in 1853, prison became the new harshest punishment. Our empirical analysis shows that the abolition of transportation differentially affected the sentencing of females to prison in a similar way as the earlier reforms. Before 1853, females were about five percentage points more likely to be sentenced to prison; after 1853, females were three percentage points less likely.

Taken together, the differential responses of juries and judges to changes in punishment severity for male and female defendants, protecting females from the harshest available punishment, provides strong support of taste-based discrimination in the form of paternalism as the underlying source of the criminal justice gender gaps. Though these ‘experiments’ had local effects on the size of the gender gap, they were, in general, not strong enough to offset the persistence of the gaps over time. Our conclusion regarding taste-based discrimination is in line with the fact that the one ‘constant’ throughout this period is the perception of females as the weaker sex. Further evidence of discrimination as the underlying source of the gender gap is provided by our analysis of the relationship between witnesses and conviction rates. Specifically, we find that conditional on observables, two prosecutorial witnesses are needed to convict females at the same rate as males with just one such witness, suggesting that prosecutors face a higher bar in terms of evidence quality for female defendants. Though such finding is consistent with the overall notion of discrimination, it does not allow us to disentangle taste-

based and statistical discrimination. The bottom line is that though we provide strong empirical and anecdotal evidence in favor of taste-based discrimination, we cannot empirically rule out that statistical discrimination also plays a role.

The remainder of the paper proceeds as follows. Section 2 provides institutional details about historical trials at the Old Bailey Central Criminal Court and describes the Proceedings trial data. Section 3 uses the Old Bailey data to illustrate changes in the (gender) composition of cases and sentencing policy from 1715 to 1900. Section 4 traces out the raw gender gaps in convictions and sentencing over time. Section 5 presents regression adjusted results, and considers the extent to which observable case characteristics can explain the gaps. Section 6 exploits the dynamic changes in sentencing policy to assess the importance of taste-based discrimination. Section 7 concludes.

2. *The Old Bailey and the Proceedings*

This paper studies the gender gap in nearly 200 years of trials from the Central Criminal Court of London and Middlesex – the Old Bailey. The details of all cases were preserved in *The Proceedings of the Old Bailey*, which were published after each court session from 1674 to 1913 (only reliable after 1715). These records have since been digitized by *The Old Bailey Proceedings Online* (Hitchcock et al., 2013) and are available as tagged xml files. This section provides a brief overview of Old Bailey trials, this unique data source, and our sample creation.

2.1. *Trials at the Old Bailey*

The jurisdiction of the Old Bailey initially included felonies in London and the surrounding county of Middlesex, but expanded in the 1830s with the addition of Essex. The definition of a felony during this period was not the same as today; offenses like pickpocketing, shoplifting, and assault, were observed at the Old Bailey, and were even capital for some of the period.

Of course, a sample of Old Bailey trials is still ‘selected’ in nature. The trials represent offenses reported by victims, arrested by police (after the 1829 introduction of the Met), charged by the magistrates, and deemed to have sufficient evidence by a Grand Jury to proceed. All of these decisions are unobserved in our data; our analysis is conditional on a case reaching ‘trial’ at the Old Bailey. It is certainly plausible (and even likely) that gender gaps exist in these earlier decisions. To the extent that they do, however, our analysis is likely to underestimate the overall criminal justice gender gap, as there is little reason to think that such earlier stage gaps do not

follow the same underlying motives as the conviction and sentencing decisions studied here.⁵

The Old Bailey trials were organized in sessions lasting at least a few days, during which a London or Middlesex specific jury decided many consecutive cases.⁶ During this period, 12 jurors were randomly chosen from a pool of potential jurors who were male, aged 21 to 60 (for most of the period), resided in England and met income/wealth qualifications.⁷ After the testimonies, the seated jury had to reach a unanimous verdict, the most common of which were acquittal, guilty, or guilty of a lesser offense. The judge decided the sentence (but the jury could recommend mercy). The main role played by women in the criminal justice system during the 1700s and 1800s was defendant. All decision makers (jurors, judges, attorneys) were male. This began to change with the Sex Disqualification (Removal) Act of 1919 (after our sample period).

2.2. Data Description: The Proceedings of the Old Bailey (1715-1900)

Initially meant to entertain readers with detailed transcripts of the most colorful cases, the Proceedings gained quasi-official (subsidized) status in the late 1700s. All trials were covered approximately equally, as the City of London demanded a ‘true, fair, and perfect narrative’.⁸ From the xml files, we extracted ‘tagged’ information identifying the unique case, session date, defendant’s name, gender and age (only available for convicted defendants after 1800), the offense as well as the verdict (plea, guilty of original or lesser charge, acquit) and sentences (death, transportation, prison, corporal, miscellaneous or no punishment). Even if defendants are charged with multiple offenses, only the main (most serious) offense is tagged in the Proceedings Online. We categorize defendants as being guilty of any charge or guilty of this original (most serious) charge versus a lesser charge. All co-defendants can be linked to the same unique case. We manually coded (i) judge, jury and juror names from 1750 to 1822 and (ii) defendant ‘criminal history’ from the 1830s onwards. The latter is in the form of untagged symbols (e.g. * or +), indicating the defendant’s previous custodial history (once, more than once, or known associate of bad character). As such a record was uncommon, we simply classify defendants as having any known history at the time of the trial.

Appendix Table 1 lists the 34 detailed offenses in the initial data and the number of observations and share of female defendants from 1715 to 1900 for each offense. To study a within-offense gender gap, we drop distinctly male (animal theft, embezzlement, mail theft,

⁵ Thus, conditioning our analysis on observed jury trials may underestimate the full extent of gender gaps in the criminal justice system. However, the fact that we look at both convictions and sentencing (and conduct robustness tests to including pleas) is already an improvement on modern-day studies of the gender gap in just sentencing.

⁶ See Bindler and Hjalmarsson’s (forthcoming) analysis of the path dependency of consecutive jury decisions.

⁷ For more details, see e.g. Beatie (1986) for the Jury Act of 1730 and Bentley (1998) for the Juries Act of 1825.

⁸ One exception is 1790-1792 when only convictions were reported. We exclude these years from our analysis.

rape, sexual assault, and sodomy) and female (infanticide) offenses; the nature of these offenses are gender specific and/or the share of female defendants is generally less than 5% (or more than 95%) in almost all sub-periods.⁹ We also drop offenses with few observations and those classified as ‘missing’ or ‘other’, which contain a wide range of not necessarily comparable offenses. We retain a final sample of 23 detailed offenses, which we categorize as property, violent, fraud or other, and 192,701 trials over this approximately 200-year period.¹⁰ As in Bindler and Hjalmarsson (2018), we also identified the capital eligibility of each offense.

Columns (1) – (3) of Table 1 provide summary statistics overall and by defendant gender for the whole period. Approximately 23% of cases had female defendants, 22% were capital, 75% were property offenses, and 13% included a guilty plea. The four main sentences were: death (7%), transportation to the Americas or Australia (30%), prison (51%), and corporal punishment (4%). The next section uses these data to illustrate dramatic changes over the sample period with respect to both the composition of cases and the predominant sanctions.

3. Criminal Justice Reforms and Institutional Changes at the Old Bailey (1715 – 1900)

3.1. Gender and Case Composition at the Old Bailey

Panel A of Figure 1 presents the annual number of cases by offense category, and provides a number of important details about trials at the Old Bailey over this time period. First, property offenses comprise the largest offense category throughout the period. During the 18th century and the latter half of the 19th century, there are generally less than 750 property trials per year, but far fewer violent, fraud and other offenses. During the 1800s, there is a large increase in the number of property crime trials from a low of 584 in 1804 to a high of 2,847 in 1843, which may due to both expanding catchment areas and the vast population growth during this period.¹¹ This is followed by a sharp decline (of almost 80%) over the next 10 years; by 1852, there were only 575 property offense trials. King (2006) suggests that this decline in property crimes can be attributed to jurisdictional changes shifting some of the less serious offenses to the lower courts or the decision-making burden for some crimes (e.g. non-violent property) away to the

⁹ There was a shift in attitudes towards female sexuality from the 18th century woman being thought to be ruled by her emotions and body to the expectation that the 19th century woman was the angel of the household. With this shift came increased concern about prostitution. In 1857, the Met Police estimated there to be 8600 prostitutes and 2825 brothels (Acton, 1857). These attitudes are unlikely to influence our analysis since (i) prostitution is not trialed at the Old Bailey, (ii) we exclude gender specific offenses, and (iii) consider the within-offense gender gap.

¹⁰ Property offenses include arson, burglary, housebreaking, larceny (combined), pickpocketing, receiving, shoplifting, stealing from master, theft from place; violent offenses include assault, manslaughter, murder, robbery (combined) and wounding; fraud offenses include coining offences, forgery and fraud; other offenses include bigamy, libel, perjury, perverting justice, return from transportation and riot.

¹¹ London’s population increased from about 750,000 in 1760 to over one million in 1801 to seven million in 1911. See <https://www.oldbaileyonline.org/static/Population-history-of-london.jsp>, viewed on August 30, 2017.

magistrates. Thus, a second take away from Panel A of Figure 1 is that the proportion of property crime trials at the Old Bailey is changing over time; it is increasing until 1843, sharply decreases in the 1840s, and is then followed by a gradual decline from 1850 to 1900 (as the number of property trials decrease and violent and fraud trials increase). One potential implication of this is that the increase in violent offense trials and the shift of minor property crimes out of the Old Bailey results in a sample of post-1850 trials at the Old Bailey that are, on average, for relatively more serious offenses than those in earlier years.

Panels B-D of Figure 1 display the annual number of male and female cases (left-hand axis) as well as the share of female defendants (right-hand axis) overall and for property and violent offenses, respectively. The share of female defendants tried at the Old Bailey declines throughout the sample period, from around 40% in the early 1700s to around 10% in 1900; this is driven by property offenses, for which an even larger decrease in the share female is seen. In contrast, the female share for violent crimes fluctuates around 10% for most of the sample period. Finally, the declining share of females for property crimes from 1750 until the mid-1800s is driven by an increasing number of male defendants more than by a decreasing number of female defendants, while the subsequent decrease in the share is driven mainly by the latter.¹²

3.2. Sentencing Reforms at the Old Bailey

This section describes the evolution of the three largest types of sanctions (death penalty, transportation, and prison) at the Old Bailey and highlights the sentencing analyzed in Section 5. Figure 2 illustrates the dramatic changes in the English penal system between 1715 and 1900. Panel A shows the share of each type of sentence for all offenses, while Panels B and C consider property and violent offenses, respectively. In each figure, the solid, long-dashed, and short-dashed black lines correspond to death, transportation, and prison sentences, respectively. We also plot the share of trials *eligible* for capital punishment in each figure. We characterize our sample period (1715-1900) by four sentencing regimes, which are clearly visible in Figure 2.

The first sentencing period includes the years prior to the American Revolution in 1776 (denoted by the first red vertical line). Prior to the Revolution, around 70% of defendants overall were annually sentenced to transportation to the Americas (Panel A), though Panels B and C

¹² This declining share of female defendants at the Old Bailey has certainly been recognized by historians studying this period; see the Old Bailey website (<https://www.oldbaileyonline.org/static/Gender.jsp#gendercrime>). There is, however, a lack of agreement about the underlying reasons for this trend in female representation at the Old Bailey, and the extent to which it is representative of other courts. Feeley and Little (1991) argue that it represents a real decline in female criminality while King (2006) argues that it is more likely to be driven by fluctuations due to times of war and peace and changing jurisdictions, especially in the second half of the 1800s. To the best of our knowledge, however, our study is the first to shed light on how the gender gap in outcomes changes with this changing composition.

show that about 80% of property versus less than 20% of violent crime trials resulted in transportation. Rather, the predominant sentence for violent crimes during this period was the death penalty (between 70 and 80%). Moreover, prior to the Revolution, prison was virtually non-existent as a sentence; the few existing prisons were primarily used as temporary holding cells for either pre-trial detention or convicted inmates awaiting transportation or death. The Revolution abruptly and unexpectedly brought a halt to this sentencing regime. The lack of a large prison infrastructure combined with the loss of the American penal colonies in 1776 led to a penal crisis in England: what should be done with convicted (property crime) offenders?

The temporary solution was to hold male convicts (females were not allowed) in the hulks of ships and put them to hard labor dredging the Thames River (*Hulks Act* of 1776). This represents the second sanction regime and corresponds to the sharp spike in ‘prison’ sentences immediately after 1776, from almost 0% to more than 40%, in Figure 2. This shock was more salient for property than violent crimes, given the relative importance of transportation.

Transportation was reinstated in 1786 with the establishment of a penal colony in Australia, marking the beginning of our third sentencing regime (see the second red vertical line in Figure 2). Transportation levels jumped sharply but never returned to the pre-Revolution levels. Rather, for the next 50 years, they stayed fairly steady with an approximate 40% (20%) sentencing share for property (violent) crimes. Transportation “beyond the seas” to Australia was particularly harsh (more-so than the Americas) given the length of the voyage, the significant chance of illness or death, and hard labor and strict discipline upon arrival. Transportation was abolished with the Penal Servitude Acts of 1853, which shifted 7-year transportation sentences into four years of penal servitude, and 1857, which abolished even longer transportation sentences. Transportation sentences drop virtually to zero after the 1853 Act (marked by the third vertical red line); there is little bite from the 1857 Act.

This third sentencing regime (from 1786 through the early 1850s) is the most dynamic. Though the share transported to Australia remains fairly stable, there are significant changes in both the use of prison and capital punishment. First, prison sentences never returned to their pre-Revolution level of 0%; rather, they remained at about 20% (10%) for property (violent) crimes. Second, the share of prison sentences began to increase in the 1820s, such that it represented more than 70% of sentences when transportation was abolished. Third, the rise of prison can largely be attributed to the abolition of capital punishment, which happened gradually (offense by offense) throughout this period, as reflected in the share of capital-eligible trials at the Old Bailey. At the beginning of the period (1786), almost 40% (100%) of the property (violent) crime trials were capital. By 1838, none of the property offenses (with the

exception of arson) and less than 10% of the violent crime trials were capital. See Bindler and Hjalmarsson (2018) for details on the year of abolition for each offense category. The abolition of capital punishment was particularly salient for violent crimes: after 1837 (when several offenses became non-capital), a sharp decrease in death sentences (from more than 60% to less than 5%) was offset by a sharp increase in prison sentences (from less than 20% to almost 70%). Thus, the third sentencing regime is characterized by transportation to Australia, the rise of prisons and the abolition of capital punishment.

This brings us to the fourth (and final) sentencing regime. By the time transportation is abolished, a new steady-state had emerged, in which incarceration was the predominant sanction for both property (more than 90%) and violent offenses (more than 80%). Of course, prison sentences for violent and property offenses could differ along the intensive margin, i.e. sentence length, but unfortunately this is not observed in the data.

The second part of the paper analyzes the effect of these sentencing reforms on the criminal justice gender gap in three natural experiments that change punishment severity, keeping in mind that each of these reforms were more/less relevant for property versus violent crimes.

3.3. Verdicts at the Old Bailey

Juries could convict the defendant of the original most serious charge, convict of a lesser charge, or acquit. Their ability to convict of a lesser charge depends on whether a suitable lesser charge existed; this was more common for property crimes, which are a function of the value of the property, than for violent crimes. Figure 3 plots the share of defendants found guilty by the jury of any offense (solid line) or the original offense (long-dashed line), and the share who plead guilty without a jury trial (short-dashed line). Panels A, B, and C show all, property and violent crime trials, respectively.

During the 1700s, the overall conviction rate remains fairly stable: approximately 60% of defendants are found guilty by the jury of any charge, with a somewhat larger (smaller) conviction rate for property (violent) crimes. The share of defendants guilty of the *original* offense is substantially smaller throughout this period: for violent offenses, it fairly steadily fluctuates around 45%, while for property crimes, there is an increasing share of defendants convicted of the original offense (from around 20% in the early 1700s to about 50% by 1760).

Figure 3 illustrates two important features of the 1800s. First, we see the introduction of pleading as a fundamental component of the judicial system. Prior to 1827, in fact, defendants were presumed guilty (hence, there was no need to plead guilty). With the introduction of a presumption of innocence, the share of defendants pleading guilty increased to about 30% by

1850 (almost 40% for property less than 10% for violent crimes). We do not focus on the gap in pleading guilty itself (a decision made by the defendant as opposed to the judge or jury). Instead, (i) we highlight that our sentencing analyses include all convictions (i.e. both jury verdicts and defendant pleas) and (ii) we test the robustness of the gender gap findings to including guilty pleas as if they were jury convictions.

The second feature of the 19th century is a general increase in jury conviction rates. For violent crimes, conviction rates jump from around 50% in the 1830s and 1840s to more than 60% (and sometimes 70%) for the remainder of the period. A similar but somewhat earlier increase is observed for property crimes in the chance of any jury conviction (from around 60% to almost 80%), which is driven by an increase in convictions of the original charge. In previous work (Bindler and Hjalmarsson, 2018), we demonstrated that these increases in *jury* conviction rates were caused by the offense-specific abolition of capital punishment (the second ‘experiment’ described above), which occurred earlier for property crimes; we also found previously that a jury’s decision to convict females was more responsive to the abolition than their decision to convict males. Bindler and Hjalmarsson (2018) thus helps to motivate our analyses of whether each sentencing reform studied here differentially affected jury and judge treatment of male and female defendants.

4. Tracing Out Raw Gender Gaps during the 18th and 19th Centuries

The first part of our analysis is descriptive: We document how the raw gender gap in jury convictions and judge sentences changed from the start of the 18th to the end of the 19th century.

4.1. Raw Gender Gaps in Convictions

For all offense categories combined, Figure 4 presents the raw share of females (gray line) and males (black line) convicted of any offense (Panel A) or the originally charged offense (Panel B). The dashed blue line plots the size of the gender gap in raw conviction rates (males-females) relative to the male mean (scaled on the right). See Appendix Figures 1 and 2 for comparable figures for property and violent crime, respectively. Figure 4 shows that, despite all of the events described above, a sizeable gender gap persists over this 200-year period, with respect to both measures of conviction. That is, in every sub-period (up to the last decade which becomes somewhat noisier), the share of males convicted of any offense is between 4 and 13 percentage points larger than for females. For conviction of the originally charged offense, males and female conviction rates both follow the same pattern over time – increasing until the mid-1800s and then decreasing again – such that the raw size of the gap ranges from 3 to 17 percentage

points up to the last decade of the sample. Relative to the male conviction rate, there is an overall decline in the relative size of the gap (for both conviction measures) through the 18th century and first half of the 19th century followed by an increase towards the end of the century.¹³

The odd-numbered columns of Appendix Table 2 present the raw gender gap (i.e. when there are no controls) in approximately 50-year periods: 1715-1750, 1751-1800, 1801-1850, and 1851-1900. Panel A shows the results for conviction of any charge (Panels B and C for convictions of the original charge and convictions or pleas, respectively). In each of these periods, the raw gender gap in the chance of a jury conviction is 11, 6, 6, and 8 percentage points, respectively (compared to 8 percentage points overall). Relative to the (male) mean conviction rates, these effect sizes translate into 17.2, 10.5, 8.7, and 11.0 percent, respectively.

4.2. Raw Gender Gaps in Sentencing

Figure 5 presents the share of convicted men and women sentenced to death, transportation, prison, and corporal punishment (from left to right), respectively;¹⁴ Panel A refers to property and Panel B to violent crimes. These figures present some interesting facts about the dynamics of the gender gap in sentencing. First, females are less likely to be sentenced to death than men while capital punishment existed. This is true for both property and violent crimes, although the gender gap appears larger for violent crimes, for which the death penalty was more salient. The capital punishment gender gap fluctuates around 5 percentage points for property crimes and 18 percentage points for violent crimes (up to 1840). Second, for property crimes, a gender gap in transportation sentences favoring females emerges in the 1740s; this spikes during the Revolution and remains somewhat larger after the emergence of the Australian penal colony. In contrast, for violent crimes, the gender gap is reversed; women are more likely to be sentenced to transportation than men through the 1700s and first decades of the 1800s. This suggests that females charged with violent crimes are ‘favored’ by being

¹³Appendix Figures 3 and 4 dig further into this pattern. Appendix Figure 3 shows the number of cases by gender and the share of female defendants for eight offenses (each with male and female defendants in each decade) covering a wide range of crime types: larceny (the most common offense), pickpocketing, receiving, theft from place, burglary, fraud, murder, and robbery. The share of female defendants declines steeply for each property crime offense, but is flatter (or even increasing) for the violent crimes of robbery and murder. Yet, even when zooming in on a specific offense (e.g. larceny) with an unambiguous decline in female representation, we see a persistently flat gap in conviction rates, as shown in Appendix Figure 4.

¹⁴Earlier versions of this paper show the gender gap in receiving the harshest punishment available for a given offense at the time. Looking at the gap in this way, one observes a very persistent gap – females were consistently less likely to be sentenced to the harshest available punishment (available upon request). But, such an aggregation hides variation across punishment categories – variation that is informative about the underlying mechanism.

transported rather than sentenced to death; consistent with this, the transportation gender gap for violent crimes disappears around the abolition of capital punishment.

Third, we turn to prison. For violent crime, little to no gender gap is seen throughout the sample period, regardless of the prevalence of prison. For property crime, however, we see that females are more likely to be sentenced to imprisonment immediately after its emergence as a sanction, i.e. after the end of the American Revolution and introduction of the transportation to Australia sentence. This pattern persists until the abolition of transportation in the 1850s at which point the gender gap in imprisonment for property crimes actually reverses; females are 10 (or more) percentage points more likely to be sentenced to prison before the abolition of transportation but about 5 percentage points less likely to be sentenced to prison afterwards. We note, however, that although the gender gap switches sign, the post 1850 gap is small relative to the mean incarceration rate, which reached more than 80% by this time.

Finally, corporal punishment was primarily used as a sanction for property crimes. Again, we see a reversal in sign of the gender gap over the sample period: Prior to the Revolution, when it was the least harsh punishment available, females were about 10 percentage points more likely to be sentenced to corporal punishment than males. With the post-Revolution emergence of prisons, females were less likely than men to receive corporal punishment.

The odd numbered columns of Appendix Table 2 summarize the persistence of the raw gender gap in sentencing, more generally, in the same four time periods (1715-1750, 1751-1800, 1801-1850, and 1851-1900). In contrast to Figure 5, we define the dependent variable as being sentenced to the harshest punishment available (for a specific offense and time period). This avoids the problem of essentially meaningless specifications where the outcome is a sanction that is not possible in these 50-year time intervals. That is, the results can be understood as a summary of the above observations: The raw gender gap in the chance of being sentenced to the harshest punishment available is 10, 14, 10 and 6 percentage points, respectively (compared to 14 percentage points overall). Relative to the (male) mean sentencing rates, these effect sizes imply that females are 18, 25, 25, and 7 percent, respectively, less likely to be sentenced to the harshest available punishment in each of the above periods. Note that both the relative and absolute gap in the sentencing rate to the harshest punishment available is smallest for the last 50-year interval. For most of the latter half of the 19th century, prison is the harshest punishment available for most offenses. That is, by this time, a number of sentencing reforms substantially decreased punishment severity.

The descriptive analysis above clearly suggests that sanction severity affects the sentencing gender gap. Our ‘experimental’ analysis returns to many of the patterns observed in this section to more formally assess the extent to which the gender gap is affected by changing sanctions.

5. Case Characteristics and Persistent Gender Gaps

5.1. Regression Adjusted Results: Controlling for Observables

Are the raw gender gaps simply an artefact of a differential distribution of offenses and case characteristics (with different conviction rates or potential punishments) by defendant gender? Do the changing gender gaps, especially in sentencing, reflect a changing composition of offenses at the courts? We formally address these explanations by estimating the gender gap in verdict and sentencing outcomes net of observable case characteristics:

$$(1) \quad Outcome_{iot} = \alpha + \beta Female_i + X_{io}\delta + \gamma_t + \varepsilon_{iot}$$

The baseline set of observables (X) includes the number of defendants, 23 detailed offense type dummies, whether the offense is capital at that time, and year fixed effects.¹⁵ The latter capture unobservable characteristics of, for instance, the justice system common to male and female defendants; our results, however, are robust to excluding the year dummies.¹⁶

Convictions: Regression Adjusted Results

We begin by estimating equation (1) for each conviction outcome by decade (these decade specific regressions omit year fixed effects). Figure 4 (C and D) presents the regression adjusted gender gap coefficient (black, solid line) and 95% confidence interval (shaded area) for jury convictions of any and the original offense. For ease of comparison, the raw gap (blue, open circles, dashed line) is also shown. For both conviction outcomes, the raw and adjusted gaps are almost identical throughout the sample period; if anything, the latter is slightly larger.¹⁷ The even-numbered columns of Appendix Table 2 confirm this conclusion: the gender gaps in convictions are not explained by observable differences in case characteristics in any of the 50-year time intervals.

¹⁵ We highlight here that there is substantial variation in the capital status of offenses. Specifically, it varies across offenses (most, though not all, were capital at some time) and over time. But, capital punishment is not abolished for all offenses at the same time; so this variation is not wiped out by year fixed effects.

¹⁶ While the most obvious differences in case characteristics, such as females committing more minor offenses, point towards an over-estimate of the gender gap, it is clearly possible for omitted variables to bias the gender gap in the other direction. We come back to the importance of potential omitted variables in our discussion of Table 3.

¹⁷ Our baseline specification is based on a linear probability model, avoiding biases associated with fixed effects in non-linear binary choice models. However, our results are robust to changing our distributional assumptions and estimating a probit instead. Results available on request.

Table 2 provides further evidence that it is not just observable differences in case characteristics that explain the underlying gender gap. Specifically, we estimate the (adjusted) gender gap separately by offense, and rank the estimates according to the share of female defendants. Focusing on the main property offenses, we see a fair bit of variation in the gender compositions: shoplifting (46%), receiving (29%), pickpocketing (27%), larceny (26%), stealing from master (23%), housebreaking (13%), robbery (13%), and burglary (7%). Significant gender gaps are seen for almost all offense categories; the corresponding gaps to the above offenses are -0.04, -0.10, -0.19, -0.07, +0.03, -0.15, -0.07, and -0.21. These results speak to the fact that observable offense characteristics (including the share of female defendants) do not explain the gender gap.

Sentencing: Regression Adjusted Results

Figure 6 shows the corresponding graphs for each sentencing category for both property and violent crimes; Appendix Figure 5 shows the results for all offenses. A similar pattern is seen for the sentencing outcomes as the conviction outcomes; the regression adjusted and raw gaps track each other very closely over time. The only exception is capital punishment for property crimes (Panel A); controlling for offense dummies matters here since there is substantial variation in capital classification. As before, Appendix Table 2 confirms these observations for our summary measure of the harshest punishment available.

5.2. Regression Adjusted Results: Controlling for a Proxy for Unobservables

Though observable differences in case characteristics do little to explain the conviction and sentencing gender gaps, this does not completely rule out that there are unobservable (to the researcher) differences in case characteristics by gender, such as the severity of a crime within a category or the quality of evidence. We can assess the extent to which unobservable differences are a likely explanation of the gaps by studying subsamples of data for which we can observe additional information.

Table 3 presents the adjusted gender gap corresponding to equation (1) for all offenses and various subsamples of the data with additional information available. Each panel corresponds to different conviction and sentencing outcomes. Column (1) presents the raw gap for the full sample period (1715-1900) and column (2) controls for the baseline set of controls (including year fixed effects). Naturally, looking at all offense categories across the entire sample period hides heterogeneity in sentencing gaps for violent and property crimes (which are subject to different sentences) and over time (as sentencing policy changed). Rather than

addressing these points (to which we return in our ‘experimental’ analyses in Section 5), the aim of this exercise is to assess the extent to which the additional controls matter (or rather do not matter) – i.e. is there selection on these ‘unobservables’? Consistent with the decade specific figures, columns (1) and (2) demonstrate that adding our baseline set of controls has little impact on the overall average gap. Columns (3) and (4) demonstrate that controlling for criminal history (only available after 1832) does not have a substantive effect on the gender gap.¹⁸

To further empirically test the importance of unobserved differences, we use the number of words per trial in the Proceedings as a proxy for unobserved characteristics for every trial in the February, May and September sessions from 1751 to 1810.¹⁹ We argue that juries may have additional information (e.g. on the number of victims, testimony length, or sensationalism) that is not observed by us, but - given that testimony is often published verbatim - likely to be captured by word count. Columns (5) and (6) of Table 3 show, however, that word count has little to no effect on the gender gap. These results suggest that it is unlikely that unobservable case differences explain a substantial portion of the gender gap, and is, in fact, consistent with minimal evidence actually being presented at trial. Feeley (1997) describes trials that were, on average, eight minutes long in the early 1800s and more consistent with the modern-day sentencing phase of a trial. While these analyses find no evidence that observables and proxies for unobservables explain any of the estimated gender gaps, we can of course not rule out the possibility that biases due to omitted variables still exist.

5.3. Within Judge and Jury Gender Gaps

One possible unobservable is a systematic assignment of females to judges and juries who are more ‘lenient’ in their treatment of all defendants. Indeed, one can observe substantial variation across juries in the share of defendants convicted (figures available upon request). To empirically test whether this plays a role, we estimate within-jury gender gaps in convictions and within-judge gender gaps in sentencing decisions for 1751 to 1822, when jury and judge information is available in the data. Yet, as seen in columns (7) and (8) of Table 3, including judge and jury fixed effects does not explain the estimated gender gaps.

¹⁸ Columns (9) and (10) of Table 3 demonstrate the robustness of the results to controlling for defendant age, which is only consistently available for convicted defendants after 1800.

¹⁹ Word count is collected for these particular years, as these years have jury identifiers. Bindler and Hjalmarsson (forthcoming) use the word count as a measure of unobservables to study path dependency in jury decision making. Regressions (shown in Appendix Table 3 of Bindler and Hjalmarsson, forthcoming) show that word count is related to observables, with generally more words for the most severe offenses, capital cases, and multiple defendants. Yet, much of the variation in word count is unexplained by observables.

5.4. *The Female as Wife and/or Mother*

The girlfriend theory, discussed in a modern day context by Starr (2015), suggests that a female is deemed not capable of being responsible of a crime; this was a formal part of the historical legal system for married couples (feme covert). Testing this explicitly in the Old Bailey data, we find that though the gap is generally larger for same name (likely married) than different name pairs of male and female co-defendants, this represents a small share of cases. As a gender gap exists also for single defendant cases, the girl-friend theory clearly cannot be the main explanation for our findings. The results are reported in Appendix Table 3.

We also argue that the importance of females as mothers is not the driving factor of the gender gaps. Using the *Old Bailey Corpus Online*, we find 2,154 utterances of the word ‘children’ (in just over 1000 trials); 39% of the utterances are by female defendants. Given that there are almost 200,000 cases, this low hit rate is by itself suggestive that the existence of children did not play a substantial role in case outcomes.²⁰ Moreover, the word ‘child’ was often used at the time in an alternative context – to be as innocent as the child unborn (many reports indeed read as if they were a plea for mercy in sentencing). One caveat of that approach is that our (best available) measure of having children is an imperfect one.

6. *Taste-based Discrimination: Evidence from Changing Sanction Regimes*

Thus far, we have demonstrated a persistent and significant gender gap in convictions and sentencing, which cannot be explained by observable and unobservable differences in case characteristics. While one can never completely rule out the possibility that there are still unobservables, this seems unlikely given the lack of sensitivity of the results to observable controls. That is, ‘selection’ on unobservables would have to be many times greater than that on observables.

This leaves *discrimination* as a remaining explanation of the gender gap. Such discrimination could, in theory, take two forms: preference or taste-based discrimination and statistical discrimination. Taste-based discrimination would imply that the all-male judiciary is less likely to convict female defendants and more likely to give them relatively lenient sentences out of a desire to protect them; such discrimination could be characterized as paternalism. Statistical discrimination, on the other hand, would occur if juries and judges, respectively, were less likely to convict or sentence females because of a general belief that females are less

²⁰ The search was conducted October 3, 2017 on <http://www1.uni-giessen.de/oldbaileycorpus/search.html>; we searched for the word ‘children’ with no constraints on speaker role or year. We note that we do not do a similar word-search for mother or feme-covert since the Old Bailey Proceedings generally do not include statements and discussions by judges or juries.

likely to commit crimes or to commit less serious crimes; in the face of perhaps imperfect information about a crime, judges and juries could use a defendant's gender as a source of additional information.

It is certainly feasible that taste-based discrimination, in the form of paternalism, explains the persistence of the gender gaps over time. In particular, this pattern lines up with what we know about societal attitudes during the 200-year period: females were persistently viewed as the 'weaker sex'. Well-known (male and female) 18th and 19th century authors make clear that females are perceived and raised to be 'inferior' to men. For instance, Reverend James Fordyce instructs women to be submissive, meek, and sensitive in his 1766 *Sermons to Young Women*, while Mary Wollstonecraft (1792) describes women as "in a state of perpetual childhood, unable to stand alone". Even in the mid-1800s, Sara Stickney Ellis (circa 1845) writes that "as women, then, the first thing of importance is to be content to be inferior to men" while John Ruskin (1865) highlights the role of men as the protector of women: "The man, in his rough work in open world, must encounter all peril and trial...But he guards the woman from all this...". Thus, given that males were deemed responsible for the welfare of females (their wives) in the home, it seems feasible that they carried this duty over to the courtroom.²¹

While this suggests that our results might plausibly be driven by taste-based discrimination, the fact that large and significant gender gaps are seen for both verdicts *and* sentencing is in contrast suggestive of a limited role for statistical discrimination. At the sentencing stage, the defendant has already been found guilty and, thus, there is no (at least, less) uncertainty left for which a signal given by gender may be useful.²²

Though the above points towards taste-based discrimination playing a relatively more important role, it is of course empirically difficult to distinguish between both types of discrimination. This section takes advantage of the dynamic changes in punishment regimes to focus on taste-based discrimination; if large changes in punishment severity differentially affect male and female conviction and/or sentencing outcomes, this is strongly suggestive of preference-based discrimination in which the all-male judiciary protects the 'weaker' sex. Moreover, this would be consistent with the main takeaways from the analysis thus far – namely that females are more (less) likely than males to be sentenced to a punishment for which there are harsher (more lenient) substitutes and that the sign of the gap changes with reforms that

²¹ The role of men as the breadwinner (Horrell and Humphries, 1995) is empirically seen in analyses (available on request) of Census data from 1851 to 1911, which show a gap in labor market participation for men and women over age 18: in 1851, 47% of females versus 95% of males ages 18-34 participated in the labor market.

²² A potential exception is if the judge feels the need to correct the jurors' behavior. Any such correction would likely run in the opposite direction of what we find and instead lead to harsher, but not more lenient sentencing.

change the relative severity of a given punishment. While we provide evidence that paternalism plays an important role in explaining the persistent gender gap in the criminal justice system, we cannot empirically rule out a parallel statistical discrimination story.

6.1. The American Revolution, the Rise of Prisons and Transportation to Australia

The first change in punishment regimes occurred with the unexpected loss of the American penal colonies during the American Revolution in 1776. We consider how the gender gaps changed during two post periods: (i) From 1776 to 1786, transportation was halted and prisons and prison hulks (for men only) were introduced. (ii) From 1787 to 1795, transportation to Australia now existed, as well as the continued and expanding use of prisons (which were harsher and more lenient, respectively, than pre-Revolution transportation).²³ To assess if and how these changes affected the gender gap, we estimate equation (2) for the years 1765 to 1795. Summary statistics are presented in Table 1. As in our earlier adjusted gender gap regressions, X includes a vector of offense dummies and the number of defendants, but it does not include capital eligibility, which did not change during this period. We consider three conviction outcomes (convicted of any offense, the original charge, and a lesser charge) and the four sentencing outcomes (death, transportation, prison, and corporal punishment).

$$(2) \quad Y_{iot} = \alpha + \beta_1 Post1_t^{76-86} + \beta_2 Post2_t^{87-95} + \beta_3 Fem_i + \beta_4 Post1_t^{76-86} * Fem_i + \beta_5 Post2_t^{87-95} * Fem_i + X_{io}\delta + \gamma_t + \varepsilon_{iot}$$

This regression provides three pieces of information: β_1 and β_2 inform us on whether juries and judges respond to the changes in expected punishment following this shift. The judge sentencing effect is, of course, partly mechanical as the availability of transportation changes; including transportation as a sentencing outcome, however, provides empirical documentation that the ‘reforms’ were implemented as we described. β_3 informs us about the gender gaps in convictions and sentencing that exist prior to the shift in sanctions, while β_4 and β_5 indicate whether these gaps significantly changed with the introduction of the new sentence regimes. Our main interest lies in the interaction terms.

A distinguishing feature of this ‘experiment’ is that the change in punishment is driven by a war in a foreign country. Thus, though the shock is exogenous to the criminal justice system, other things may change concurrently with the war. This implies that we can only capture the reduced form effect of the Revolution. Unfortunately, we do not have a large enough sample of non-transportation eligible offenses to estimate a difference-in-differences model to

²³ Transportation (in name only) was re-introduced in 1781, as there was no sustainable penal colony yet.

control for potential other changes during the war.²⁴ We note, however, this limitation may be less relevant for the post-war period (as opposed to during the war). We also highlight here that we have previously studied this natural experiment (Bindler and Hjalmarsson, 2018) to learn about the effect of expected punishment on jury verdicts, finding that the temporary halt of transportation reduced the chance of conviction for non-capital offenses. Here, we expand our previous analyses to consider alternative measures of conviction and sentencing outcomes, and, importantly, whether there are differential effects by gender.

The Shift in Punishment Regimes and Gender Gaps in Convictions

Table 4 presents the jury conviction and judge sentencing results in columns (1)–(3) and (4)–(7), respectively; Panels A, B, and C present results for all, property, and violent offenses, respectively. Standard errors are clustered by offense.²⁵ There are a number of interesting and relevant findings. First, consistent with the gender gaps in conviction shown earlier, females are less likely to be convicted of any charge (9 percentage points) and the original charge (12 percentage points) but more likely to be convicted of a lesser charge (4 percentage points). This is largely driven by property crime, but similar effects are seen for violent crime. Are females differentially affected by the reforms? Yes, with the evolution of prison as a sentencing option, i.e. in the second post period, females are differentially more likely to be convicted of the original charge (about +6 percentage points) rather than a lesser charge (-4 percentage points). This is also seen for both property and violent crimes.

Thus, the gender gap in convictions gets smaller as the *minimum* expected punishment decreases from pre-Revolution transportation to prison. Though the gap decreases, however, we note that it is not completely eliminated: females are 12 percentage points less likely to be convicted of the original charge pre-Revolution but 6 percentage points less likely after. Moreover, the gap in convictions of any charge does not change. Finally, an important caveat is that we just study a short period after the shock, and do not test the persistence of this effect – this is hard to do given all of the other dynamic historical changes.

²⁴ Of course, the inability to implement a differences-in-differences design may raise questions about the causal interpretation of these pre-post analyses. The robustness of the results, however, to controlling for a linear trend (in the spirit of a regression discontinuity design) provides support for such an interpretation. In particular, Appendix Table 7 shows that the main coefficients of interest (for female plus the interaction terms) are robust to (i) not including year fixed effects nor any trend, (ii) including a linear trend instead of year fixed effects and (iii) including offense-specific linear trends.

²⁵ The overall samples includes 22 offenses, which is clearly a borderline (but generally) acceptable number of clusters. Clustering by offense for the property (8 offenses) and violent (5 offenses) crime samples is potentially more problematic. However, Appendix Tables 4-6 show the robustness of our results to alternatively using robust standard errors. If anything, using the clustered standard errors as the baseline is the more conservative choice.

The Shift in Punishment Regimes and Gender Gaps in Sentencing

In terms of sentencing, we see a clear decrease in transportation during the temporary halt that persists into the post period; that is, consistent with historical descriptions, transportation becomes unavailable during the war. During the temporary halt, transportation eligible offenses are shifted into prison/prison hulks and corporal punishment (approximately equally). During the second post-period, a significant increase in sentences to prison, corporal punishment, and miscellaneous or no punishment is observed (the latter two categories are not shown). Most importantly for any conclusion regarding taste-based discrimination, females are disproportionately more likely (14 percentage points) to be sentenced to prison during both the temporary halt and the second post period. Of course, one might question whether this represents preferences or simply capacity constraints (i.e. there may have been fewer positions for females in Australia and females are excluded from the hulks). However, seeing a similarly disproportionate shift for females away from corporal punishment (which is not capacity constrained) speaks towards preference-based discrimination being the dominant channel. In particular, column (7) shows that females were 5 percentage points more likely to be sentenced to corporal punishment before the Revolution, but 8 percentage points (0.05–0.13) less likely after the introduction of (harsh) transportation to Australia.²⁶

In summary, we find empirical evidence that both judge and jury decisions are differentially affected by the changes in relative punishment severity after the American Revolution, suggesting that they are ‘protecting’ female defendants from the harsher punishment regimes.

6.2. The Abolition of Capital Punishment

The next shift in punishment severity came with the offense-specific abolition of capital punishment between 1807 (pickpocketing) and 1856 (arson). In Bindler and Hjalmarsson (2018), using a slightly different sample, we found that abolishing the death penalty increased the chance of any conviction by more than 7 percentage points (driven by violent and sex offenses) and the chance of conviction of the original charge by 16 percentage points (driven by property offenses). To avoid the death penalty, jurors had to acquit (or convict but recommend mercy) for violent offenses, but could more easily convict of lesser (non-capital) charges for property offenses. Moreover, we found that the abolition differentially affected the conviction of male and female defendants charged with violent offenses.

²⁶ This is consistent with earlier descriptions of Figure 6 and confirms that the change is in fact significant.

We use the same difference-in-differences approach and sample years (1803-1871) to replicate and expand the Bindler and Hjalmarsson (2018) analysis for the sample of non-gender specific offenses in this paper and a wider range of outcomes, including conviction of lesser offenses and judge sentences. See columns (6) and (7) of Table 1 for summary statistics.

As in Bindler and Hjalmarsson (2018), we identify the year of abolition for each offense and classify offenses as never, always or once capital eligible.²⁷ Intuitively, the always and never capital offenses serve as ‘control’ offenses while those for which the death penalty was abolished are ‘treated’ (with a staggered treatment). We compare the change in conviction and sentencing rates in the years surrounding reforms for defendants charged with treated offenses to the change for control offenses, thereby controlling for other changes in society that affect both treated and control offenses. We use this framework to assess whether these reforms differentially affect male and female defendants, and the corresponding gender gaps.

Panels A-C, respectively, of Figure 7 illustrate the intuition of the difference-in-differences model by plotting the annual share of death, transportation, and prison sentences separately for the treatment group (black) and control group (gray) as well as for females (solid lines) and males (dashed lines). Since treatment occurs in multiple years, we center the figure on the crime-specific year of reform for the treatment offenses (see the vertical line); for the control offenses, we use a weighted average of years corresponding to the share of reforms occurring in a given year. Panel A shows that males charged with treated offenses were sentenced to death at a higher rate than females prior to abolition (40% versus 25%) and that the share of death sentences for both males and females drops sharply to zero at the time of the reform; this mechanical elimination of the gender gap also provides clear evidence that the death penalty abolition was implemented, as required by the law. Panel B shows that the use of transportation for both treated and control offenses is increasing somewhat in the years before the abolition and then starts to decline around the time of capital abolition (as prisons expand). Yet, for treated offenses, there is a sharp increase in transportation sentences around the abolition for both females and males. How does the gender gap change around the reform? It appears that the increase in transportation for males in the treatment group exceeds the increase for females in the treatment group; we test this more formally in our regression analysis. Finally, Panel C illustrates that prison levels for control offenses were generally higher than treated offenses (before and after the reform and for males and females). The abolition sharply increased the share of male and female offenders sentenced to incarceration; we rely again on

²⁷ See Appendix Table 2 in Bindler and Hjalmarsson (2018).

the formal regressions to assess whether this affected the size of the gender gap in incarceration.

Equation (3) presents the formal difference-in-differences model for each conviction and sentencing outcome Y for defendant i charged with offense o in year t .

$$(3) \quad Y_{iot} = \alpha + \beta_1 Noncapital_{ot} + \beta_2 Fem_i + \beta_3 Fem_i * Noncapital_{ot} + \alpha_o + \alpha_t + \mathbf{X}_i \delta + \epsilon_{iot}$$

Noncapital is an indicator equal to one for offense-year combinations for which the offense is not capital eligible. As offense fixed effects are a fundamental part of the identification (it is the offense that determines treatment status), we explicitly separate them out from X (the number of defendants). Year fixed effects (α_t) capture other potential shocks and reforms that affect all offenses. β_1 , β_2 , and β_3 inform us, respectively, on how the abolition of capital punishment affected judicial behavior overall, whether there was a gender gap in the pre-period, and whether this gap changed with the abolition of capital punishment.^{28,29}

To assign a causal interpretation to these coefficients requires a number of identifying assumptions. The first is the usual parallel trends assumption, i.e. that the changes in conviction rates and sentencing outcomes for treatment offenses would have been the same as for control offenses in the absence of the reforms. Parallel pre-trends in Figure 7 already suggest this is plausible. Other assumptions are that the timing of the offense-specific abolition was random and that the composition of cases and associated evidence did not change after the reforms. Bindler and Hjalmarsson (2018) provides extensive empirical support of these assumptions: (i) results are robust to offense-specific time trends, (ii) there are no significant ‘leads’ in event-study specifications, (iii) there is no significant change in the chance of pleading guilty nor in observable criminal behavior, and (iv) the quality of evidence, measured using keyword searches for police, evidence and witness, does not significantly increase. We refer the reader to this previous paper for the respective results, and further discussions of the identifying assumptions.

The Abolition of Capital Punishment and Gender Gaps in Convictions

Table 5 presents the main results. We begin with the conviction outcomes in columns (1)-(3). Consistent with our earlier work, the abolition of capital punishment increases the chance

²⁸ Standard errors are again clustered by offense. Bindler and Hjalmarsson (2018) shows that the conclusions are unchanged when alternatively using one-way and two-way clustering, block bootstrap as well as wild t-bootstrap. Appendix Tables 4-6 show our results with robust instead of clustered standard errors.

²⁹ As shown in Bindler and Hjalmarsson (2018), the capital punishment analyses are robust to excluding never capital and always capital offenses, respectively.

of any conviction by 19 percentage points for violent crimes and increases (decreases) the chance of conviction of the original charge (a lesser charge) by 16 (15) percentage points for property crimes. More importantly, how are the pre-existing gender gaps affected? For violent offenses, the 12 percentage point gap in the chance of conviction is almost completely offset by the abolition of capital punishment. Similarly, for property crimes, the gender gap in conviction of a lesser charge (about 5 percentages points) is offset by the reform, but there is no significant change in the gender gap for the chance of being convicted of the original or any charge. Again, the lack of an overall effect on conviction squares with the persistence of the conviction gender gap throughout the sample, which is dominated by property offenses.

The Abolition of Capital Punishment and Gender Gaps in Sentencing

In terms of sentencing, the abolition of capital punishment completely eliminates the previously existing gender gap in death sentences (column (4)); again, this mechanical effect simply shows that the abolition of the death penalty was implemented in practice as prescribed by the law. The more interesting question is whether a gender gap arises in other (substitute) sentences. The interaction coefficients for transportation, prison, and corporal punishments in columns (5)-(7) are negative, but generally insignificant. The only statistically significant estimate (10% level) is found for transportation for violent crimes. Thus, while we see some response of juries to abolishing capital punishment that depend on defendant gender, we do not find much evidence that judges differentially respond to the new sanctions. It may be that we simply cannot observe changes in judge behavior, as there is no flexibility/nuance in sentencing at the extensive margin. However, as we cannot observe the ‘intensive’ margin, we cannot rule out that abolishing capital punishment increases the gender gap in prison sentence lengths or the likelihood of getting a pardon.

In summary, we find statistical evidence that the decrease in punishment severity due to the death penalty abolition differentially affected males and females: jurors and judges disproportionately protected female defendants *before* the abolition.

6.3. The Abolition of Transportation

The abolition of capital punishment was eventually followed by the abolition of transportation. Since, as described above, the 1853 Act ended transportation for most offenses while the 1857 Act had little additional bite, we focus on the former. Did the 1853 abolition of transportation differentially effect male and female conviction rates and sentences? As with the temporary halt and introduction of transportation to Australia (Section 6.1), we are limited to a pre-post

specification (with a differential treatment effect for males and females) due to a lack of a suitable control group. We focus on the sample years 1845 to 1860 and estimate equation (4):

$$(4) \quad Y_{iot} = \alpha + \beta_1 Post_t^{1853} + \beta_2 Fem_i + \beta_3 Post_t^{1853} * Fem_i + X_{io}\delta + \gamma_t + \varepsilon_{iot}$$

As before, Y denotes the respective outcome (verdict, sentence) for defendant i charged with offense o in year t . $Post^{1853}$ is a dummy equal to one in 1853 and later and X is a vector of observable controls (number of defendants, capital eligibility and offense dummies). Year fixed effects (γ_t) capture shocks affecting all offenses. β_1 , β_2 , and β_3 inform us, respectively, on how the 1853 abolition of transportation affected the respective outcome overall, whether there was a gender gap in the pre-period, and whether this gap changed with the reform.

The abolition of transportation falls into a dynamic period of reforms, which poses some caveats to our analysis. First, the 1853 reform follows a period of reforms abolishing capital punishment offense by offense (arson was not until 1856), and it is possible that the (persistent) effects last into our sample period. Second, just before the 1853 abolition of transportation, we see a change in the composition of offenses, as cases are shifted out of the Old Bailey and into lower courts. Finally, the abolition of transportation is the last of a series of reforms lowering the severity of punishments; at the same time, while the 1853 reform constitutes the *de jure* abolition of transportation, *de facto* there was already a large decrease in transportation sentences (for property offenses) leading up to the reform (see Section 3.2). Given these potential confounders, our main aim is to assess whether there was a differential impact of the transportation abolition on male and female defendants and not its overall impact per se.³⁰

The Abolition of Transportation and Gender Gaps in Convictions

The results are shown in Table 6. In terms of conviction, the abolition only had a significant impact for violent offenses: it increased (decreased) the chance of conviction of the original (lesser) charge. More to the point, however, there is no differential impact by gender for either property or violent offenses: the point estimate on the interaction term is never significantly different from zero and is in fact close to zero in most cases.

The Abolition of Transportation and Gender Gaps in Sentencing

³⁰ Moreover, the main results are again – as in the case of the temporary halt of transportation – robust to controlling for linear time trends (i.e. in the spirit of a regression discontinuity design around the reform). See Appendix Table 8 for the results when we (i) do not include year fixed effects nor a trend, (ii) include a linear trend instead of year fixed effects and (iii) include offense-specific linear trends. As before, the main coefficients of interest (for female and for the interaction terms) prove to be robust to the different specification choices.

Turning to the sentencing outcomes in columns (4) to (7), we first note that (reassuringly) there is no impact on death sentences; this suggests that our empirical specification is capable of disentangling the effect of the abolition of transportation from the abolition of capital punishment, which was nearing its end in our sample period. The decrease in transportation in column (5) for both property and violent crimes confirms that transportation was abolished in practice. Further, females were less likely to be sentenced to transportation before 1853; this gap is mechanically eliminated upon the abolition of transportation. As seen in Section 3, the change in the sentencing regime led to an increase in prison sentences; column (6) of Table 6 indicates that there was a 25 percentage point increase in the chance of prison after 1853. Most interestingly, females were about five percentage points more likely to be sentenced to prison before 1853, but three percentage points less likely after 1853 (0.0473-0.0782). That is, the abolition of transportation differentially affected the sentencing of females to prison in a manner consistent with our earlier findings – namely that a female favoring gap is introduced in a sanction when that sanction is the harshest punishment available, i.e. as the *relative* severity of prison as a punishment increased. These results are not seen for violent crimes, for which there is a relatively low share of offenses; however, such gender gaps could have been introduced in sentencing margins, e.g. sentence length, which we cannot observe.

6.4. Further Evidence of Discrimination: Differential Quality of Evidence

The results from the analyses of sentencing regime shifts suggest that taste-based discrimination (or paternalism) plays an important role in explaining the observed gender gaps. Such forms of discrimination would imply that juries are willing to acquit females with a relatively high degree of evidence against them, as any disutility due to the potential release of a criminal is offset by their positive taste for protecting females. In other words, preference-based discrimination should raise the quality of evidence threshold needed to convict a female relative to a male. We note that a similar implication could be found, however, in a statistical discrimination scenario – i.e. where juries require a higher level of evidence to believe that a particular female is guilty, given potential beliefs about females being less criminal on average.

Measuring quality of evidence – in even a modern-day context – is notoriously difficult. We consider one potential input into a measure of quality of evidence; we manually coded the number of prosecution witnesses in each February, May and September session from 1800 to 1810 (almost 3500 cases). Given that we do not (and cannot) evaluate the character of or quality of each witness, this is clearly an imperfect measure. Yet, we believe it to be a valid component: the more witnesses brought by the prosecution, the stronger the case. In fact, almost all cases

(99.7%) with no prosecution witnesses are acquitted. Do prosecutors get a lower return on witness testimony for female versus male defendants?

Table 7 presents the regression adjusted gender gap for this sample using the baseline specification in column (1), with controls for prosecution witness in columns (2) and (3), and interactions with the number of witnesses and defendant gender in columns (4) and (5). Similar to our main sample, females are almost nine percentage points less likely to be convicted and this is not affected by controlling for the number of witnesses. But, male defendants with one witness against them are convicted, on average, 63% of the time while females with one witness are convicted just 50% (0.63-0.13) of the time. When there are two or more witnesses, males are convicted 74% of the time versus 67% for females. In other words, prosecutors actually need (at least) two witnesses for female defendants to achieve approximately the same conviction rate as that for male defendants with one witness. This average lower return to prosecutorial witness testimony for female versus male defendants can reflect the preferences or beliefs ('discrimination') of the jury, who hold the witnesses to a higher standard, or even the preferences of the witnesses themselves, who may behave differently depending on the defendant's gender. Thus, these results provide further evidence consistent with a story of discrimination, though they are not conclusive about whether it is taste-based or statistical discrimination.

7. Conclusion

This paper finds a criminal justice gender gap in convictions and sentencing that (i) favors females, (ii) persists through almost 200 years, (iii) is seen across offense categories, and (iv) is not explained by observable and proxies for unobservable case characteristics. Though we cannot empirically rule out that statistical discrimination plays a role, we do provide empirical evidence consistent with a taste-based discrimination mechanism. Consistent with paternalism, there is a differential response of the all-male jury in convicting and the all-male judiciary in sentencing male and female defendants when faced with changing punishment regimes. That is, a role that was played by men in society at this time appears to spill over into the courtroom.³¹

Specifically, gender gaps in jury convictions tend to get smaller when punishment severity decreases. However, though these local changes in jury behavior inform us on the importance of taste-based discrimination as an underlying mechanism, they are generally not

³¹ A natural follow-up question is how the gap was affected when female representation was introduced to the judiciary. Anwar et al. (2019) find that the addition of females to the Old Bailey jury pool in 1921 impacted conviction rates for more 'female salient' cases, including sex offenses, male versus female victim cases, and female defendants charged with other (including abortion) offenses.

large enough to eliminate these persistent gender gaps in conviction. Judges persistently react to these reforms by imposing a sentencing gap on whichever sentence becomes (in relative terms) the harshest punishment available after a sentencing reform.

This conclusion of a taste-based discrimination mechanism is in line with the modern-day findings of Schanzenbach (2005) and Philippe (forthcoming) that the presence of female judges decreases the gender gap, which they interpret as evidence of a paternalistic bias of male judges. However, they contrast the conclusions of Butcher et al. (2017), based on a rank order test, that tastes for gender discrimination are unlikely to drive the gender gap in sentencing.

Are the decisions, and underlying attitudes towards females, of judges and jurors of this historical period informative (externally valid) to that of their contemporary counterparts? During our 200-year sample, there is a persistent, explicit bias towards women as the weaker sex. This bias is seen in the courtroom (to the benefit of women). While today's writings do not exhibit the same explicit bias as in the 18th and 19th centuries, these beliefs may still be prevalent and carried over to the courtroom as an implicit bias. As gender gaps are certainly still seen in contemporary criminal justice outcomes, (male and female) agents of the justice system today may be unconsciously and unintentionally discriminating (Bertrand et al., 2005) on the basis of gender. Despite all of the ground gained by women, this perception may not have completely disappeared – even if it is less spoken.

References

- Acton, William (1857) *Prostitution Considered in its Moral, Social, and Sanitary Aspects, in London and Other Large Cities*, London.
- Altonji, Joseph and Rebecca Blank (1999), “Race and Gender in the Labor Market”, in O. Ashenfelter and D. Card, (eds), *Handbook of Labor Economics Volume 3c* Elsevier Science B.V.: 3144-3259
- Anwar, Shamena, Patrick Bayer, and Randi Hjalmarsson (2019) “A Jury of Her Peers: The Impact of the First Female Jurors on Criminal Verdicts”, *Economic Journal*, 129(618): 603-650.
- Beattie, J.M. (1986) *Crime and the Courts in England 1660-1800*, Princeton University Press, pp 663ff.
- Bentley, David (1998) *English Criminal Justice in the Nineteenth Century*, The Hambledon Press, London, pp 318ff.
- Bertrand, Marianne, Dolly Chugh, and Sendhil Mullainathan (2005) “Implicit Discrimination”, *AEA Papers and Proceedings*, 95(2): 94-98.
- Bindler, Anna and Randi Hjalmarsson (2018) “How Punishment Severity Affects Jury Verdicts: Evidence from Two Natural Experiments”, *American Economic Journal: Economic Policy*, 10(4): 36-78.
- Bindler, Anna and Randi Hjalmarsson (forthcoming) “Path Dependency in Jury Decision-

Making”, *Journal of the European Economic Association*.

Blau, Francine and Lawrence Kahn (2017) “The Gender Wage Gap: Extent, Trends, and Explanations,” *Journal of Economic Literature*, 55(3): 789-865.

Bodenhorn, Howard (2009) “Criminal Sentencing in 19th-century Pennsylvania”, *Explorations in Economic History*, 46: 287-298.

Butcher, Kristin, Kyung H. Park and Anne Morrison Piehl (2017) “Comparing Apples to Oranges: Differences in Women’s and Men’s Incarceration and Sentencing Outcomes”, *Journal of Labor Economics*, 35(S1): S201-S234.

Feeley, Malcolm (1997) “Legal Complexity and the Transformation of the Criminal Process: The Origins of Plea Bargaining,” *Israel Law Review*, 31: 183-222.

Feeley, Malcolm and Little, Deborah (1991) “The Vanishing Female: The Decline of Women in the Criminal Process, 1687-1912”, *Law and Society Review*, 25: 719-57.

Fordyce, James (1766) *Sermons to Young Women in Two Volumes*. The Sixth Edition, Volume I., D. Payne, London.

Hitchcock, Tim, Robert Shoemaker, Clive Emsley, Sharon Howard, Jamie McLaughlin, et al. “The Old Bailey Proceedings Online, 1674-1913.” www.oldbaileyonline.org (version 7.1, April 2013).

Horrell, Sara and Jane Humphries (1995) “Women’s Labour Force Participation and the Transition to the Male-Breadwinner Family, 1790-1865,” *The Economic History Review*, 48(1): 89-117.

King, Peter (2006), *Crime and Law in England, 1750-1840*, Cambridge

Mustard, David B. (2001) “Racial, Ethnic, and Gender Disparities in Sentencing: Evidence from the U.S. Federal Courts,” *Journal of Law and Economics*, 44: 285-314.

Nicholas, Stephen and Deborah Oxley (1993) “The Living Standards of Women during the Industrial Revolution, 1795-1820”, *The Economic History Review*, 46(4): 723-749.

Philippe, Arnaud (forthcoming) “Gender Disparities in Criminal Justice,” *Economica*.

Ruskin, John (1865) *Sesame and Lilies: Two Lectures*, Second Edition, Smith, Elder, and Co., London.

Schanzenbach, Max M. (2005) “Racial and Gender Disparities in Prison Sentences: The Effect of District-Level Judicial Demographics,” *Journal of Legal Studies*, 34: 57-92.

Shermer, Lauren O’Neill, and Brian Johnson (2010) “Criminal Prosecutions: Examining Prosecutorial Discretion and Charge Reductions in U.S. Federal District Courts,” *Justice Quarterly*, 27: 394-430.

Sorensen, Todd, Supriya Sarnikar, and Ronald Oaxaca (2012) “Race and Gender Differences under Federal Sentencing Guidelines,” *American Economic Review Papers and Proceedings*, 102(3): 56-60.

Spohn, Cassia and Jeffrey Spears (1997) “Gender and Case Processing Decisions,” *Women and Criminal Justice*, 8: 29-59.

Starr, Sonja (2015) “Estimating Gender Disparities in Federal Criminal Cases,” *American Law and Economics Review*, 17(1): 127-519.

Stickney Ellis, Sara (1845) *The Daughters of England*. Fisher, Son, and Co., London and Parts.

Stolzenberg, Lisa and Stewart D’Alessio (2004) “Sex Differences in the Likelihood of Arrest,”

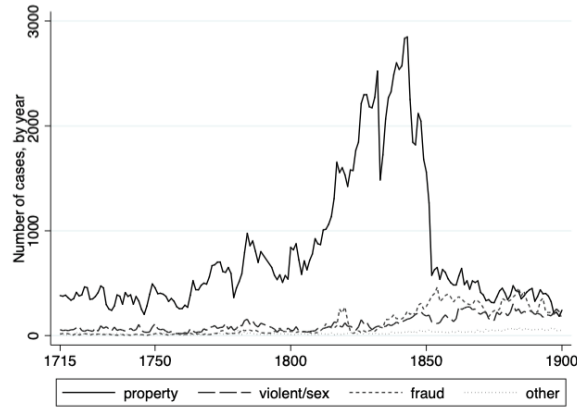
Journal of Criminal Justice, 32: 443-54.

Vickers, Chris (2016) "Socioeconomic Status and Judicial Disparities in England and Wales, 1870-1910", *Explorations in Economic History*, 61: 32-53.

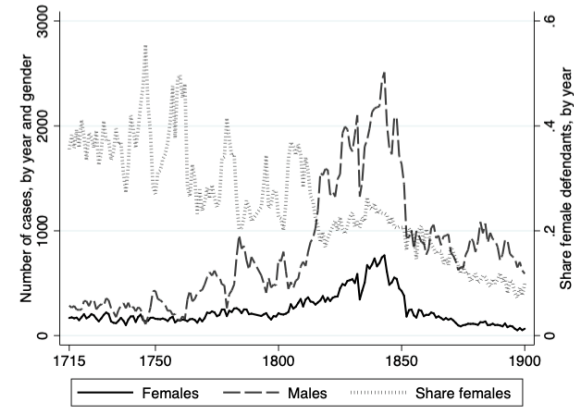
Wollstonecraft, Mary (1792) *A Vindication of the Rights of Woman: with Strictures on political and Moral Subjects*. Printed at Boston, by Peter Edes for Thomas and Andrews, Faust's statue, no. 45, Newbury-street, MDCCXCII. Pp. 452.

Figure 1. Number of Cases by Broad Offense Category and Gender

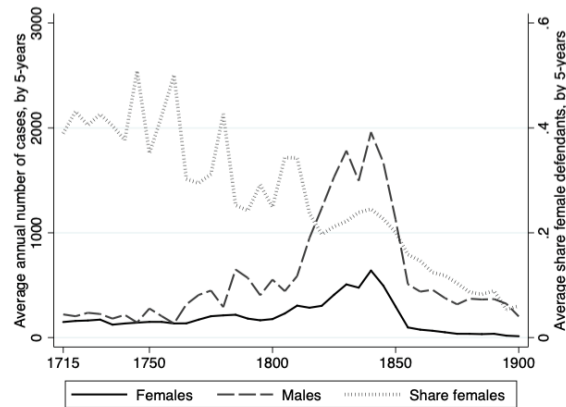
A. Total number of cases



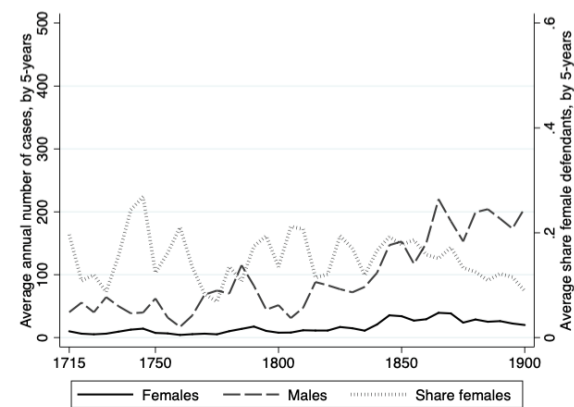
B. Number of cases by gender



C. Property offense by gender



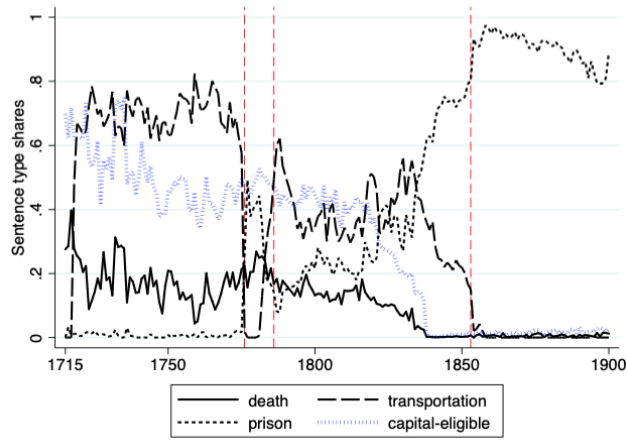
D. Violent offenses by gender



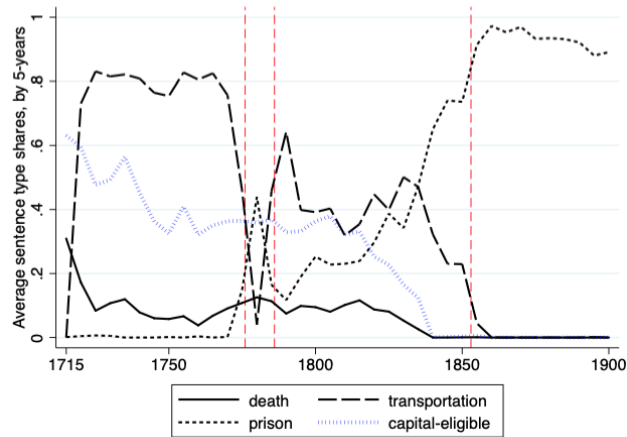
NOTE – Figure A shows the annual number of cases by offense category, Figure B the annual number of cases by year and gender (left axis) as well as the annual share of female defendants (right axis). Figures C and D show the average annual number of cases and the average share of female defendants (both averaged over 5-year intervals) for property and violent offenses, respectively. SOURCE - *The Old Bailey Proceedings Online* and own calculations.

Figure 2. Sentencing Reforms and Trends in Sentence Types

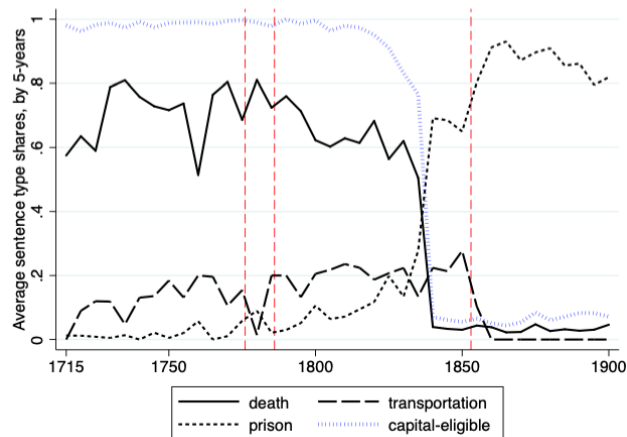
A. All offenses



B. Property offenses



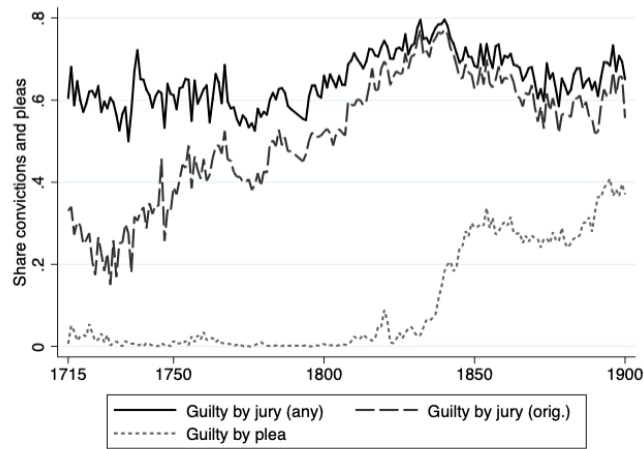
C. Violent offenses



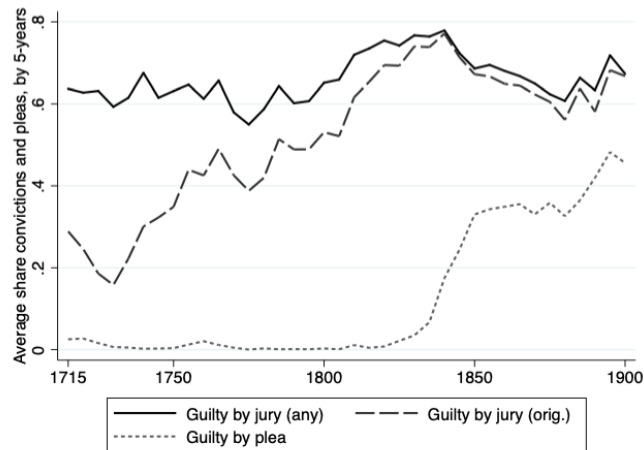
NOTE – Figure A shows the annual share of cases by sentence type (conditional on conviction) as well as the share of capital-eligible cases for all offense types. Figures B and C show the average annual shares (averaged over 5-year intervals) separately for property and violent offenses, respectively. The red vertical lines mark the years during the halt of transportation due to the American Revolution (1776-1786) as well as the abolition of transportation in 1853 (1857). Offense-specific reforms abolishing capital punishment started in 1808. SOURCE - *The Old Bailey Proceedings Online* and own calculations.

Figure 3. Trends in Conviction Rates and Pleas

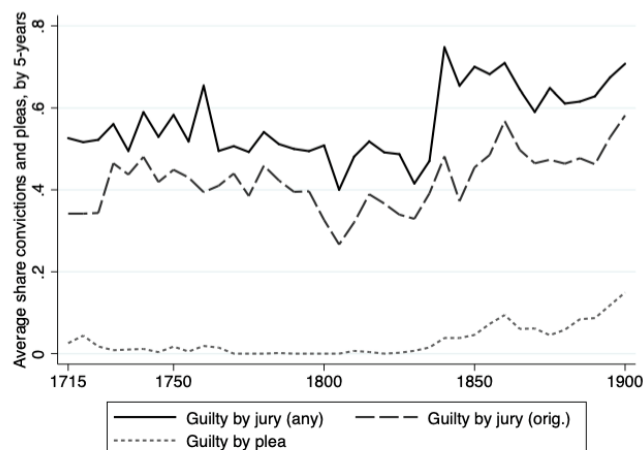
A. All offenses



B. Property offenses



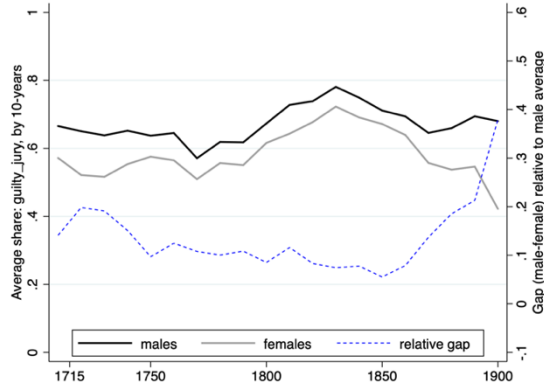
C. Violent offenses



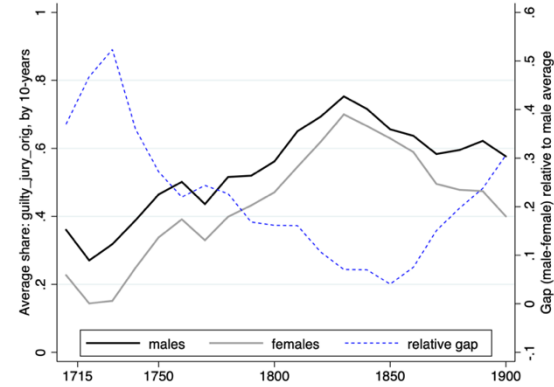
NOTE – Figure A shows the annual share of convictions and pleas for all offense types. Figures B and C show the average annual shares (averaged over 5-year intervals) separately for property and violent offenses, respectively. SOURCE - *The Old Bailey Proceedings Online* and own calculations.

Figure 4. Raw and Adjusted Gender Gaps in Conviction (All Offenses)

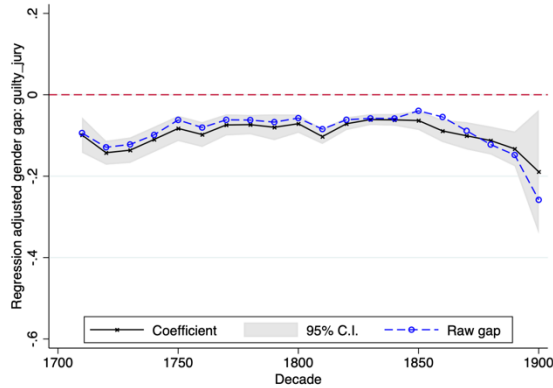
A. Raw gap in jury convictions, any offense



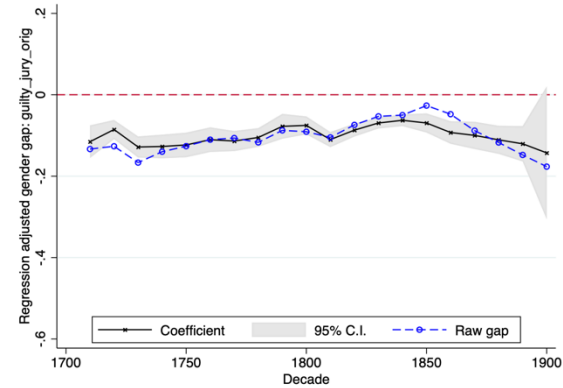
B. Raw gap in jury convictions, original offense



C. Adjusted gap in jury convictions, any offense



D. Adjusted gap in jury convictions, original offense

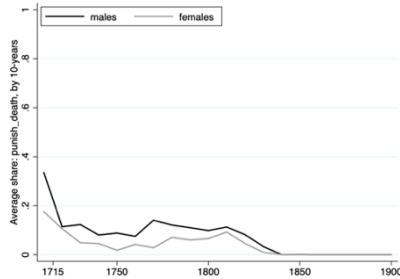


NOTE – Figure A (B) shows the average annual share of jury convictions of any (original) offense by gender (averaged by decade) as well as the raw gender gap relative to the male mean. Figures C and D plot the corresponding adjusted gender gap and 95% confidence interval as well as the raw gender gap. The adjusted gap corresponds to the estimated coefficients from equation (1) including controls for gender, number of defendants, capital eligibility and offense fixed effects. SOURCE - *The Old Bailey Proceedings Online* and own calculations.

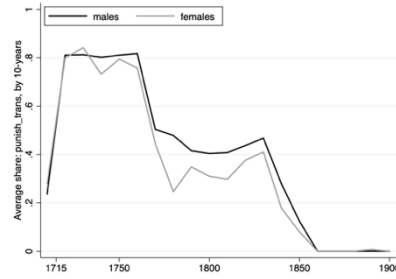
Figure 5. Raw Gender Gaps in Sentencing for Property and Violent Offenses

Panel A. Property offenses

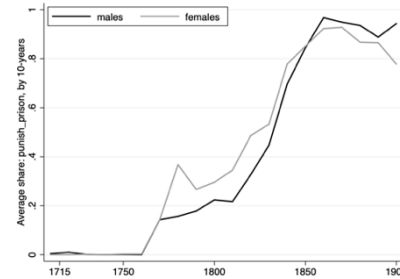
A1. Capital punishment



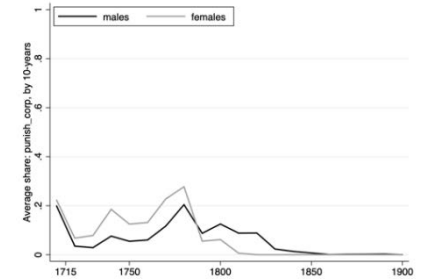
A2. Transportation



A3. Prison

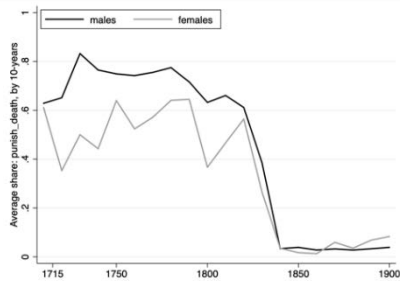


A4. Corporal punishment

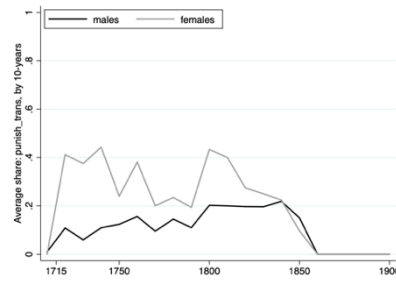


Panel B. Violent offenses

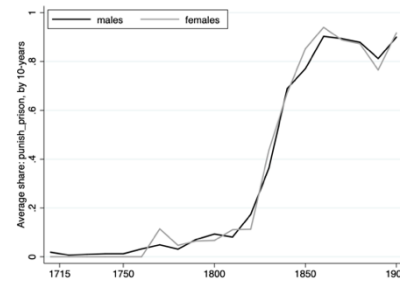
B1. Capital punishment



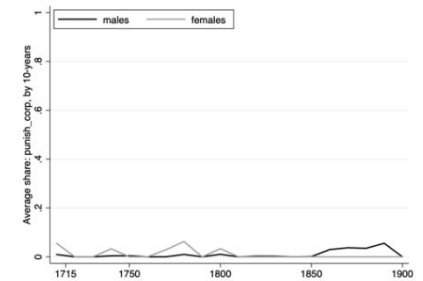
B2. Transportation



B3. Prison



B4. Corporal punishment

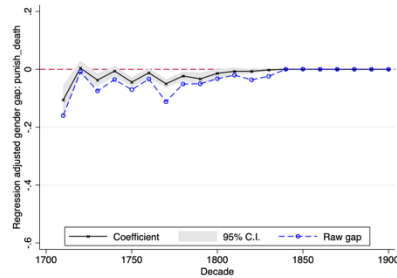


NOTE – The figures show the average share of cases for males (black line) and females (gray line) sentenced to capital punishment (A1/B1), transportation (A2/B2), prison (A3/B3) and corporal punishment (A4/B4), each by gender and for property (A) and violent (B) offenses, respectively (averaged by decade and conditional on conviction by jury or plea). Offense-specific reforms abolishing capital punishment started in 1808, transportation was step-wise abolished in 1853-57. SOURCE - *The Old Bailey Proceedings Online* and own calculations.

Figure 6. Adjusted Gender Gaps in Sentencing for Property and Violent Offenses

Panel A. Property offenses

A1. Capital punishment



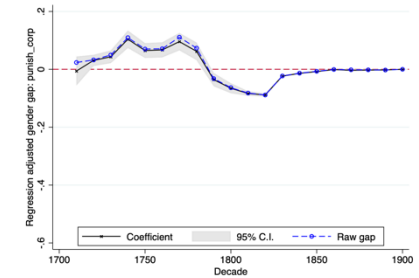
A2. Transportation



A3. Prison

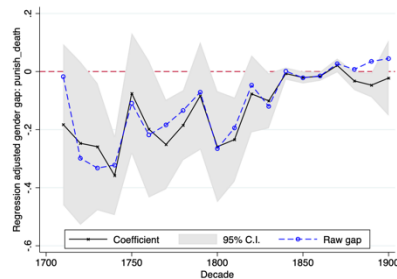


A4. Corporal punishment

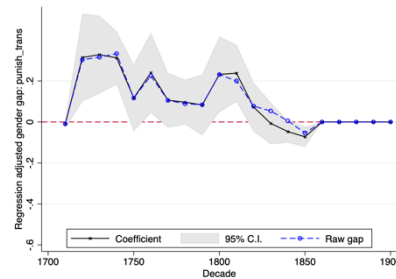


Panel B. Violent offenses

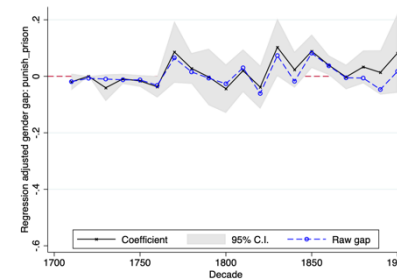
B1. Capital punishment



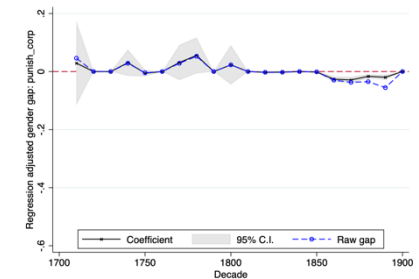
B2. Transportation



B3. Prison



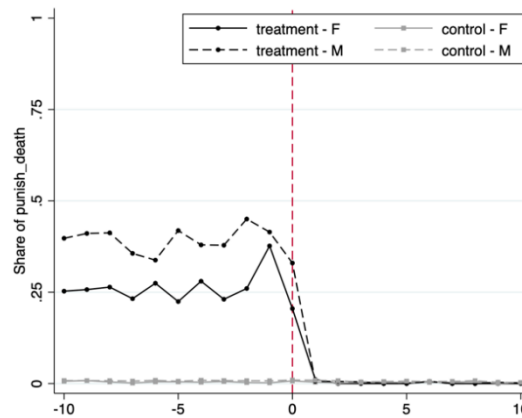
B4. Corporal punishment



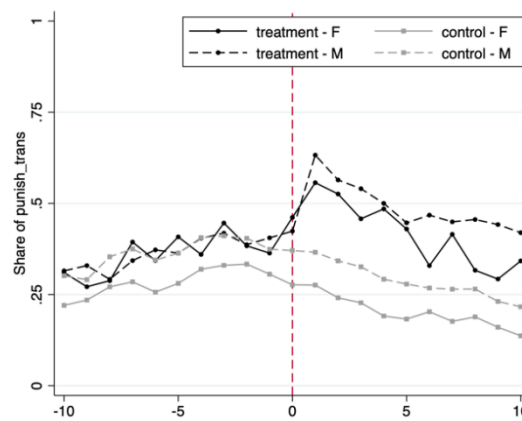
NOTE – The figures show the adjusted gender gaps (black solid line) and 95% confidence intervals for four different sentencing types, separately for property (A) and violent (B) offenses. The adjusted gap corresponds to the estimated coefficients from equation (1) including controls for gender, number of defendants, capital eligibility and offense fixed effects. The raw gender gap (blue, open circles, dashed line) is also shown. Figures A1/B1 show the results for capital punishment, A2/B2 for transportation, A3/B3 for prison and A4/B4 for corporal punishment (by decade and conditional on conviction). Offense-specific reforms abolishing capital punishment started in 1808, transportation was step-wise abolished in 1853-57. SOURCE - *The Old Bailey Proceedings Online* and own calculations.

Figure 7. Graphical Depiction of Capital Punishment Abolition

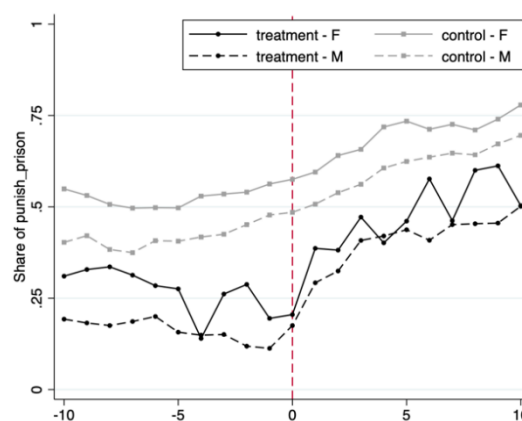
A. Capital punishment



B. Transportation



C. Prison



NOTE – The figure shows the annual share of convicted cases in the treatment (black) and control (grey) group that were sentenced to death (Panel A), transportation (Panel B) or prison (Panel C) in the 10 years before and after the assigned treatment year, separately by gender (solid: female; dashed: male). The treatment group contains those offenses for which capital punishment is abolished, and the control group those that are never or always capital eligible. The vertical line marks the offense specific year of abolition of capital punishment for offenses in the treatment group. The reform period for the control group is a weighted average by the share of reforms in each treatment year. SOURCE - *The Old Bailey Proceedings Online* and own calculations

Table 1. Summary Statistics

		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
			<u>1715-1900</u>		<u>1765-1795</u>		<u>1803-1971</u>		<u>1845-1860</u>	
		all	male	female	male	female	male	female	male	female
<u>Trial Characteristics</u>	N	192,701	148,314	43,736	23,503	9,346	91,598	25,756	41,336	10,503
capital	mean	0.215	0.205	0.247	0.466	0.408	0.145	0.151	0.008	0.010
property off.	mean	0.751	0.731	0.820	0.811	0.917	0.791	0.821	0.684	0.687
violent off.	mean	0.107	0.118	0.068	0.116	0.042	0.083	0.058	0.117	0.095
sex off.	mean	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
fraud off.	mean	0.118	0.124	0.096	0.046	0.030	0.109	0.111	0.176	0.204
special off.	mean	0.003	0.004	0.001	0.012	0.001	0.002	0.000	0.002	0.000
other off.	mean	0.021	0.022	0.016	0.015	0.009	0.015	0.009	0.022	0.014
plea	mean	0.129	0.146	0.073	0.005	0.003	0.151	0.098	0.273	0.188
guilty (plea + guilty by jury)	mean	0.722	0.744	0.649	0.622	0.555	0.777	0.709	0.801	0.738
<u>Verdict Conditional on Jury Trial</u>	N	167,796	126,709	40,546	23,378	9,318	77,783	23,226	30,047	8,532
guilty by jury	mean	0.681	0.700	0.621	0.620	0.554	0.738	0.678	0.726	0.677
guilty by jury, original charge	mean	0.586	0.615	0.498	0.500	0.391	0.689	0.624	0.682	0.644
guilty by jury, lesser charge	mean	0.099	0.088	0.131	0.379	0.446	0.262	0.322	0.273	0.322
acquitted by jury	mean	0.318	0.299	0.378	0.127	0.178	0.050	0.057	0.044	0.033
guilty with recommendation for mercy	mean	0.060	0.059	0.062	0.026	0.019	0.079	0.093	0.080	0.100
<u>Sentences Conditional on Conviction or Plea</u>	N	135,568	107,543	27,624	14,348	5,057	69,650	17,827	32,165	7,509
death penalty	mean	0.069	0.075	0.047	0.196	0.076	0.051	0.034	0.003	0.002
transportation	mean	0.301	0.295	0.326	0.463	0.432	0.305	0.257	0.166	0.115
prison	mean	0.511	0.519	0.480	0.136	0.206	0.554	0.613	0.796	0.836
corporal punishment	mean	0.043	0.042	0.046	0.110	0.157	0.035	0.003	0.008	0.000
harshest punishment	mean	0.523	0.552	0.408	0.549	0.407	0.461	0.352	0.480	0.365

NOTE – The table shows summary statistics for the variables in the whole sample and subsamples by time period. Where not otherwise specified, the mean of the variable is shown. SOURCE - *The Old Bailey Proceedings Online* and own calculations.

Table 2. Offenses with High and Low Share Female Defendants (1715-1900)

(1) off.	(2) N	(3) share female	(4) coeff.	(5) s.e.	(6) relative gap
Jury conviction of any charge					
Shoplifting	3,460	0.455	-0.0351**	(0.0148)	-0.043
Receiving	8,510	0.293	-0.0952***	(0.0119)	-0.193
Coining offenses	8,713	0.282	-0.0726***	(0.0100)	-0.092
Pickpocketing	14,021	0.27	-0.1917***	(0.0090)	-0.243
Larceny	66,244	0.264	-0.0662***	(0.0040)	-0.089
Theft from place	14,628	0.261	-0.0776***	(0.0089)	-0.108
Arson	463	0.248	-0.0650	(0.0620)	-0.151
Stealing from Master	7,322	0.233	0.0301**	(0.0121)	0.044
Perjury	1,110	0.205	-0.0146	(0.0403)	-0.046
Murder	2,095	0.179	-0.1700***	(0.0291)	-0.291
Manslaughter	1,822	0.178	-0.0688**	(0.0294)	-0.183
Bigamy	1,017	0.171	-0.1777***	(0.0393)	-0.255
Wounding	4,262	0.161	-0.0412**	(0.0188)	-0.058
Perverting justice	488	0.159	-0.0415	(0.0726)	-0.069
Housebreaking	3,662	0.131	-0.1548***	(0.0234)	-0.203
Robbery	10,615	0.131	-0.0663***	(0.0143)	-0.109
Libel	321	0.104	-0.0647	(0.1121)	-0.123
Fraud	4,123	0.0906	-0.0585**	(0.0260)	-0.089
Forgery	3,613	0.0827	-0.0795***	(0.0291)	-0.137
Burglary	9,452	0.0733	-0.2064***	(0.0180)	-0.294
Assault	773	0.0674	-0.0597	(0.0736)	-0.099
Return from transportation	313	0.0514	0.0326	(0.1353)	0.041
Riot	228	0.048	0.2431*	(0.1387)	0.484

NOTE – The table shows the estimated coefficients when regressing the outcome (jury conviction for any charge) on a dummy for female defendants, by detailed offense. Column (1) lists the offense; column (2) shows the number of observations (1715-1900), column (3) the share females, column (4) the estimated coefficient, column (5) the standard error and column (6) the gap relative to the male mean. The regressions include observable controls (see Table 3) without offense fixed effects. Robust standard errors are shown in parentheses. *** p<0.01, ** p<0.05, *p<0.1

Table 3. Regression Adjusted Gender Gaps: Observables and Unobservables

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
<i>Sample:</i>	1715-1900		>1832		1751-1810		1751-1822		>=1800	
<i>Spec.:</i>	no ctrls.	obs. ctrls.	obs. ctrls.	+crim. hist.	obs. ctrls.	+wordcount ctrl.	obs. ctrls.	+jury/judge f.e.	obs. ctrls.	+age ctrl.
Panel A. Jury conviction of any charge										
Female	-0.0790*** (0.0176)	-0.0803*** (0.0155)	-0.0693*** (0.0155)	-0.0544*** (0.0131)	-0.0827*** (0.0191)	-0.0763*** (0.0171)	-0.0878*** (0.0164)	-0.0818*** (0.0181)		
Panel B. Jury conviction of original charge										
Female	-0.1170*** (0.0210)	-0.0979*** (0.0145)	-0.0700*** (0.0165)	-0.0552*** (0.0139)	-0.1075*** (0.0085)	-0.1008*** (0.0069)	-0.1082*** (0.0130)	-0.1026*** (0.0145)		
Panel C. Jury conviction of any charge or guilty by plea										
Female	-0.0950*** (0.0186)	-0.0835*** (0.0157)	-0.0750*** (0.0141)	-0.0653*** (0.0122)	-0.0835*** (0.0188)	-0.0773*** (0.0169)	-0.0883*** (0.0170)	-0.0822*** (0.0184)		
Panel D. Sentenced to death										
Female	-0.0274 (0.0181)	-0.0234 (0.0142)	-0.0041 (0.0024)	-0.0042 (0.0025)	-0.0319 (0.0263)	-0.0245 (0.0224)	-0.0266 (0.0201)	-0.0261 (0.0186)	-0.0107 (0.0064)	-0.0111 (0.0069)
Panel E. Sentenced to transportation										
Female	0.0309 (0.0316)	-0.0398 (0.0286)	-0.0424** (0.0180)	-0.0369** (0.0165)	-0.0906*** (0.0293)	-0.0913*** (0.0265)	-0.0842*** (0.0280)	-0.0849*** (0.0240)	-0.0489** (0.0224)	-0.0483** (0.0195)
Panel F. Sentenced to prison										
Female	-0.0392 (0.0255)	0.0468** (0.0182)	0.0282 (0.0215)	0.0240 (0.0197)	0.0579*** (0.0057)	0.0553*** (0.0058)	0.0843*** (0.0076)	0.0849*** (0.0064)	0.0567** (0.0248)	0.0476** (0.0184)
Panel G. Sentenced to corporal punishment										
Female	0.0042 (0.0048)	-0.0097** (0.0042)	-0.0097*** (0.0028)	-0.0097*** (0.0028)	0.0454*** (0.0066)	0.0439*** (0.0064)	-0.0099 (0.0059)	-0.0105* (0.0050)	-0.0321*** (0.0110)	-0.0283*** (0.0087)
Observations										
Panels A-B	167 255	167 255	72 861	72 861	13 372	13 372	56 777	54 711		
Panel C	192 050	192 050	95 886	95 886	13 415	13 415	57 657	55 251		
Panels D-G	135 167	135 167	72 231	72 231	8 126	8 126	36 787	35 648	108 497	101 667

NOTE –The table shows the estimated coefficients when regressing the outcome (verdicts in panels A-C and sentences conditional on conviction in panels D to G) on a dummy for female defendants, corresponding to estimating equation (1). Columns (1) and (2) show the results for the whole sample period with/without observable controls (number defendants, capital eligibility, detailed offense category and year fixed effects). Column (4) adds a control variable for criminal history, column (6) jury and judge fixed effects (without year fixed effects), column (8) a control for the trial report word count (scaled by 100) and column (10) a quadratic function in age (only available for convicted defendants). To ease comparison, columns (3), (5), (7) and (9) show the corresponding results with the baseline observable controls for each available sample period, respectively. Robust standard errors are clustered at the offense level and shown in parentheses. *** p<0.01, ** p<0.05, *p<0.1

Table 4. Shift in Punishment Regimes - American Revolution (1765-1795)

Outcome:	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	any charge	<i>Jury conviction</i> original charge	lesser charge	death	<i>Judge sentence</i> transportation	prison	corporal
Panel A. All offenses							
Post 76-86	-0.0293 (0.0332)	-0.0001 (0.0140)	-0.0393 (0.0301)	0.0075 (0.0334)	-0.2367*** (0.0724)	0.1156*** (0.0373)	0.1360*** (0.0416)
Post 87-95	-0.0503* (0.0285)	-0.0151 (0.0245)	-0.0435 (0.0438)	-0.0042 (0.0357)	-0.4024*** (0.0698)	0.1107*** (0.0281)	0.0436*** (0.0108)
Female	-0.0899*** (0.0180)	-0.1190*** (0.0111)	0.0356** (0.0130)	-0.0297 (0.0272)	-0.0004 (0.0617)	-0.0242* (0.0123)	0.0529*** (0.0177)
Post 76-86 x Fem.	0.0180 (0.0158)	0.0025 (0.0143)	0.0138 (0.0158)	-0.0297 (0.0206)	-0.1960*** (0.0572)	0.1358*** (0.0399)	0.0728*** (0.0136)
Post 87-95 x Fem.	0.0169 (0.0180)	0.0579*** (0.0127)	-0.0391** (0.0145)	0.0248 (0.0228)	-0.0395 (0.0326)	0.1417*** (0.0390)	-0.1311*** (0.0439)
N	21,164	21,164	21,164	12,145	12,145	12,145	12,145
Cluster	22	22	22	22	22	22	22
Panel B. Property offenses							
Post 76-86	-0.0134 (0.0353)	0.0216*** (0.0057)	-0.0461 (0.0366)	0.0030 (0.0376)	-0.2624*** (0.0748)	0.1289*** (0.0359)	0.1551*** (0.0384)
Post 87-95	-0.0670** (0.0283)	-0.0221 (0.0249)	-0.0534 (0.0519)	0.0161 (0.0421)	-0.4183*** (0.0694)	0.1168*** (0.0275)	0.0433*** (0.0093)
Female	-0.0874*** (0.0187)	-0.1077*** (0.0101)	0.0264 (0.0141)	-0.0233 (0.0246)	-0.0374 (0.0493)	-0.0109 (0.0080)	0.0632*** (0.0153)
Post 76-86 x Fem.	0.0145 (0.0170)	-0.0072 (0.0173)	0.0211 (0.0184)	-0.0237 (0.0241)	-0.1470** (0.0461)	0.1093** (0.0334)	0.0610*** (0.0147)
Post 87-95 x Fem.	0.0089 (0.0189)	0.0433*** (0.0118)	-0.0328 (0.0181)	0.0204 (0.0232)	-0.0137 (0.0295)	0.1448*** (0.0410)	-0.1500*** (0.0406)
N	17,575	17,575	17,575	10,313	10,313	10,313	10,313
Cluster	8	8	8	8	8	8	8
Panel C. Violent offenses							
Post 76-86	0.0274 (0.1258)	0.0043 (0.0138)	0.0303 (0.1393)	0.0241 (0.0593)	-0.1098*** (0.0184)	0.0453 (0.0291)	0.0051 (0.0052)
Post 87-95	-0.0076 (0.1319)	0.0165 (0.0094)	-0.0213 (0.1561)	0.0116 (0.0511)	-0.3329*** (0.0620)	0.1301 (0.0636)	-0.0003 (0.0031)
Female	-0.0173 (0.0497)	-0.1412** (0.0346)	0.1296 (0.0707)	-0.2509 (0.1292)	0.2383 (0.1220)	-0.0153 (0.0284)	0.0026 (0.0042)
Post 76-86 x Fem.	-0.0229 (0.0391)	0.0200 (0.0464)	-0.0503 (0.0621)	0.0153 (0.1388)	-0.1336 (0.1049)	0.0823 (0.0756)	0.0801*** (0.0035)
Post 87-95 x Fem.	0.0138 (0.0576)	0.0928** (0.0291)	-0.0625 (0.0747)	0.1790 (0.1220)	-0.1678 (0.1078)	-0.0001 (0.0074)	0.0010 (0.0018)
N	2,281	2,281	2,281	1,138	1,138	1,138	1,138
Cluster	5	5	5	5	5	5	5

NOTE – The table shows the results corresponding to estimating equation (2) for the sample period 1765-1795. The dependent variable is indicated at the top of each column (verdicts in columns (1) to (3) and sentences in columns (4) to (7)). The regressions control for the number of defendants, detailed offense category and year fixed effects. Panel A shows the results for all offenses, Panel B for property and Panel C for violent offenses, respectively. Robust standard errors are clustered at the offense level and shown in parentheses. *** p<0.01, ** p<0.05, *p<0.1

Table 5. The Abolition of Capital Punishment (1803-1871)

Outcome:	(1) any charge	(2) <i>Jury conviction</i> original charge	(3) lesser charge	(4) death	(5) <i>Judge sentence</i> transportation	(6) prison	(7) corporal
Panel A. All offenses							
Noncapital	0.0809* (0.0439)	0.1693*** (0.0377)	-0.0932** (0.0390)	-0.4138*** (0.0792)	0.1941** (0.0733)	0.0927** (0.0352)	0.0554*** (0.0139)
Female	-0.0649*** (0.0161)	-0.1064*** (0.0178)	0.0530*** (0.0169)	-0.1121*** (0.0390)	-0.0052 (0.0440)	0.0977*** (0.0227)	-0.0217*** (0.0049)
Noncapital x Female	-0.0032 (0.0219)	0.0320 (0.0254)	-0.0465** (0.0179)	0.1165*** (0.0402)	-0.0548 (0.0533)	-0.0353 (0.0336)	-0.0156 (0.0122)
N	101,009	101,009	101,009	87,477	87,477	87,477	87,477
Cluster	23	23	23	23	23	23	23
Panel B. Property offenses							
Noncapital	0.0204 (0.0259)	0.1630*** (0.0448)	-0.1502*** (0.0372)	-0.3770*** (0.0962)	0.2035* (0.0900)	0.0505* (0.0269)	0.0493*** (0.0114)
Female	-0.0758*** (0.0154)	-0.1192*** (0.0188)	0.0563*** (0.0156)	-0.1110* (0.0530)	-0.0117 (0.0539)	0.1080*** (0.0245)	-0.0281*** (0.0050)
Noncapital x Female	0.0017 (0.0242)	0.0390 (0.0299)	-0.0500** (0.0174)	0.1172* (0.0542)	-0.0567 (0.0632)	-0.0359 (0.0366)	-0.0163 (0.0111)
N	81,504	81,504	81,504	70,776	70,776	70,776	70,776
Cluster	9	9	9	9	9	9	9
Panel C. Violent offenses							
Noncapital	0.1857* (0.0807)	0.1222 (0.0807)	0.0793 (0.1367)	-0.8181*** (0.0531)	-0.0081 (0.0370)	0.6954*** (0.0951)	0.0216*** (0.0043)
Female	-0.1217** (0.0414)	-0.1391*** (0.0162)	0.0264 (0.0429)	-0.1471** (0.0391)	0.0974 (0.0544)	0.0239** (0.0065)	0.0019 (0.0019)
Noncapital x Female	0.1158** (0.0329)	0.1150*** (0.0126)	-0.0091 (0.0415)	0.1321** (0.0416)	-0.1165* (0.0521)	0.0174 (0.0148)	-0.0156 (0.0118)
N	8,702	8,702	8,702	5,578	5,578	5,578	5,578
Cluster	5	5	5	5	5	5	5

NOTE – The table shows the results corresponding to estimating equation (3) for the sample period 1803-1871. The dependent variable is indicated at the top of each column (verdicts in columns (1) to (3) and sentences in columns (4) to (7)). The regressions control for the number of defendants, detailed offense category and year fixed effects. Panel A shows the results for all offenses, Panel B for property and Panel C for violent offenses, respectively. Robust standard errors are clustered at the offense level and shown in parentheses. *** p<0.01, ** p<0.05, *p<0.1

Table 6. The Abolition of Transportation (1845-1860)

Outcome:	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	<i>Jury conviction</i>			<i>Judge sentence</i>			
	any charge	original charge	lesser charge	death	transportation	prison	corporal
Panel A. All offenses							
Post 1853	-0.0089 (0.0315)	-0.0024 (0.0390)	-0.0065 (0.0160)	0.0028 (0.0022)	-0.2689*** (0.0488)	0.2463*** (0.0504)	-0.0022 (0.0021)
Female	-0.0529** (0.0237)	-0.0541** (0.0243)	0.0012 (0.0047)	-0.0001 (0.0013)	-0.0705*** (0.0167)	0.0473** (0.0210)	-0.0100*** (0.0020)
Post 1853 x Female	-0.0307 (0.0207)	-0.0335 (0.0246)	0.0029 (0.0090)	-0.0022 (0.0027)	0.0781*** (0.0272)	-0.0782*** (0.0272)	0.0098*** (0.0016)
N	18,591	18,591	18,591	19,875	19,875	19,875	19,875
Cluster	23	23	23	23	23	23	23
Panel B. Property offenses							
Post 1853	-0.0322 (0.0421)	-0.0348 (0.0519)	0.0025 (0.0189)	-0.0002 (0.0002)	-0.3022*** (0.0641)	0.2832*** (0.0627)	-0.0049* (0.0024)
Female	-0.0571 (0.0319)	-0.0609 (0.0334)	0.0038 (0.0050)	-0.0002 (0.0002)	-0.0702*** (0.0206)	0.0473 (0.0269)	-0.0128*** (0.0010)
Post 1853 x Female	-0.0398 (0.0281)	-0.0482 (0.0375)	0.0085 (0.0186)	0.0019 (0.0023)	0.0707* (0.0349)	-0.0867** (0.0355)	0.0112*** (0.0012)
N	11,866	11,866	11,866	13,380	13,380	13,380	13,380
Cluster	9	9	9	9	9	9	9
Panel C. Violent offenses							
Post 1853	0.0846 (0.0615)	0.1744*** (0.0350)	-0.0898** (0.0319)	0.0148 (0.0159)	-0.2264 (0.1311)	0.1873 (0.1617)	-0.0007 (0.0005)
Female	-0.0517 (0.0539)	-0.0432 (0.0414)	-0.0085 (0.0364)	-0.0001 (0.0158)	-0.0611* (0.0235)	0.0454** (0.0134)	-0.0011 (0.0009)
Post 1853 x Female	-0.0087 (0.0238)	-0.0145 (0.0512)	0.0058 (0.0394)	-0.0233 (0.0243)	-0.0285 (0.0512)	0.0470 (0.0537)	0.0007 (0.0005)
N	2,546	2,546	2,546	1,887	1,887	1,887	1,887
Cluster	5	5	5	5	5	5	5

NOTE – The table shows the results corresponding to estimating equation (4) for the sample period 1845-1860. The dependent variable is indicated at the top of each column (verdicts in columns (1) to (3) and sentences in columns (4) to (7)). The regressions control for the number of defendants, capital eligibility, detailed offense category and year fixed effects. Panel A shows the results for all offenses, Panel B for property and Panel C for violent offenses, respectively. Robust standard errors are clustered at the offense level and shown in parentheses. *** p<0.01, ** p<0.05, *p<0.1

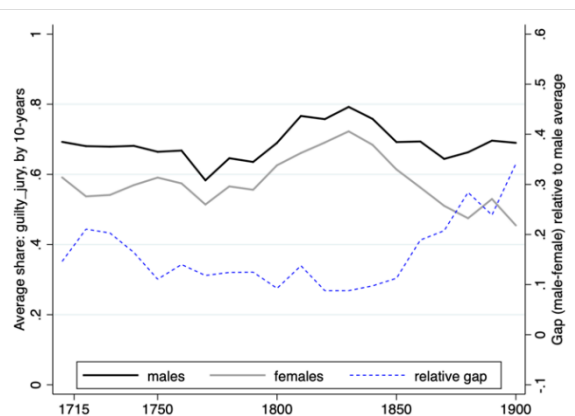
Table 7. Prosecution Witnesses (1800-1810)

Specification:	(1) baseline	(2) + control	(3) + control	(4) + interaction	(5) + interaction
<i>Jury conviction of any charge</i>					
Female	-0.0851*** (0.0172)	-0.0852*** (0.0181)	-0.0835*** (0.0178)	-0.0104 (0.0130)	-0.0118 (0.0126)
Any witness		0.7067*** (0.0507)		0.7304*** (0.0494)	
Female x any witness				-0.0821*** (0.0204)	
One witness			0.5905*** (0.0766)		0.6325*** (0.0833)
Two or more witnesses			0.7187*** (0.0509)		0.7393*** (0.0488)
Female x one witness					-0.1309*** (0.0304)
Female x two or more witnesses					-0.0719*** (0.0226)
Observations	3,480	3,480	3,480	3,480	3,480

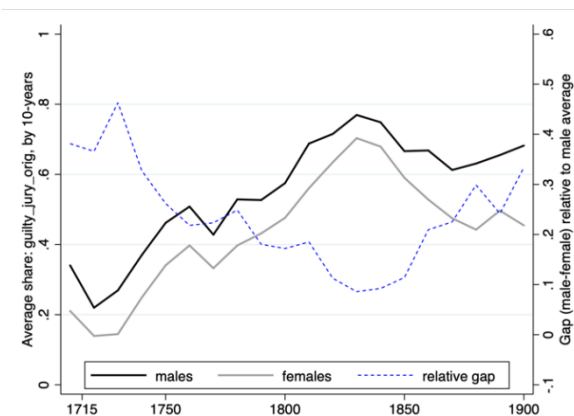
NOTE – The table shows the estimated coefficients when regressing the outcome (jury conviction of any charge) on a dummy for female defendants using the sample including information on the number of prosecution witnesses. The sample includes the February, May and September sessions for the years 1800 to 1810. Column (1) includes observable controls (see Table 3), column (2) and (3) control for the presence of any or one and two and more witnesses, respectively, and column (4) and (5) allows for interactions of that measure with the female dummy variable. Robust standard errors are clustered at the offense level and shown in parentheses. *** p<0.01, ** p<0.05, *p<0.1

Appendix Figure 1. Raw and Adjusted Gender Gaps in Conviction (Property Crimes)

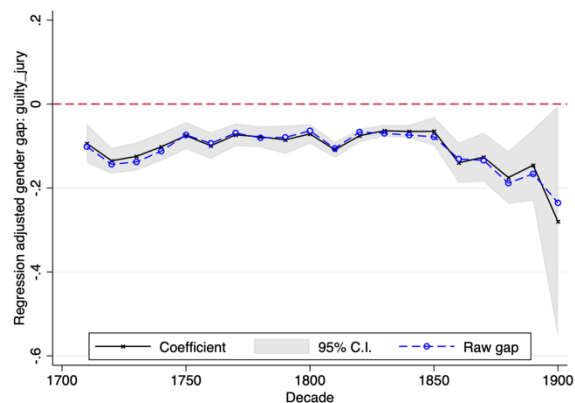
A. Raw gap in jury convictions, any offense



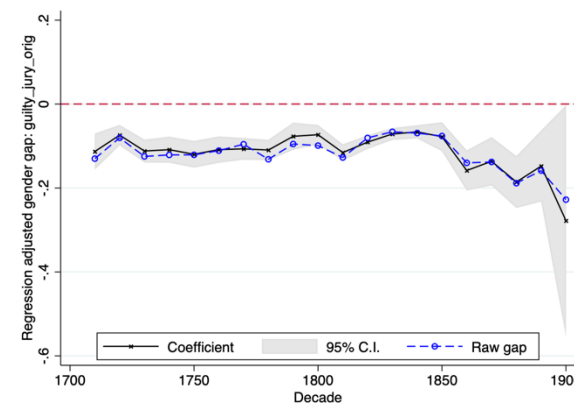
B. Raw gap in jury convictions, original offense



C. Adjusted gap in jury convictions, any offense



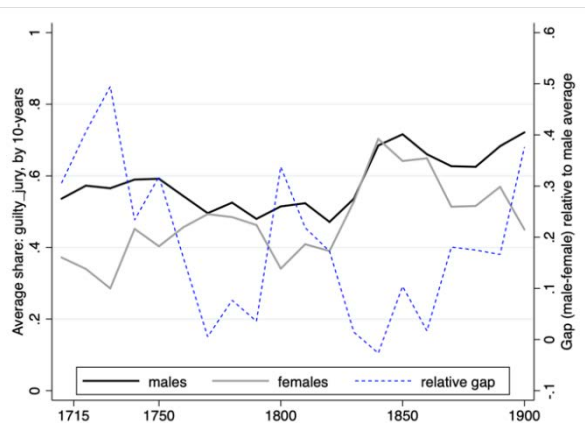
D. Adjusted gap in jury convictions, original offense



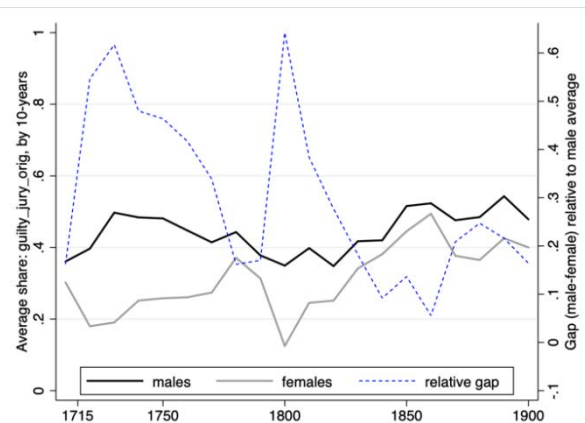
NOTE – Figure A (B) shows the average annual share of jury convictions of any (original) offense by gender (averaged by decade) as well as the raw gender gap relative to the male mean. Figures C and D plot the corresponding adjusted gender gap and 95% confidence interval as well as the raw gender gap. The adjusted gap corresponds to the estimated coefficients from equation (1) including controls for gender, number of defendants, capital eligibility and offense fixed effects. The sample is restricted to property crimes only. SOURCE - *The Old Bailey Proceedings Online* and own calculations.

Appendix Figure 2. Raw and Adjusted Gender Gaps in Conviction (Violent Crimes)

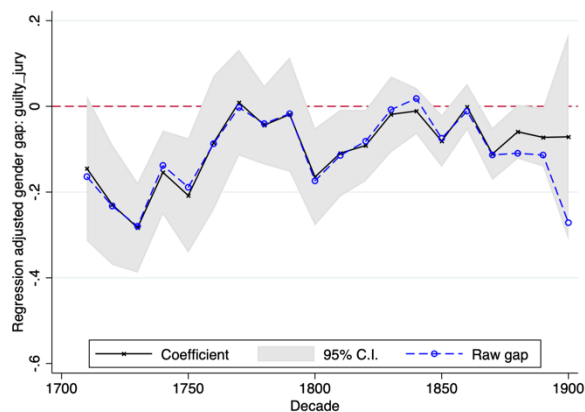
A. Raw gap in jury convictions, any offense



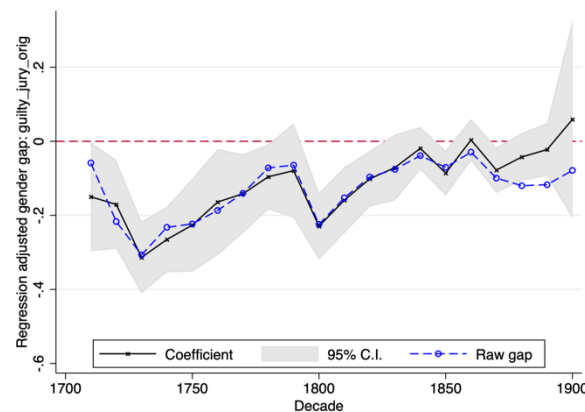
B. Raw gap in jury convictions, original offense



C. Adjusted gap in jury convictions, any offense

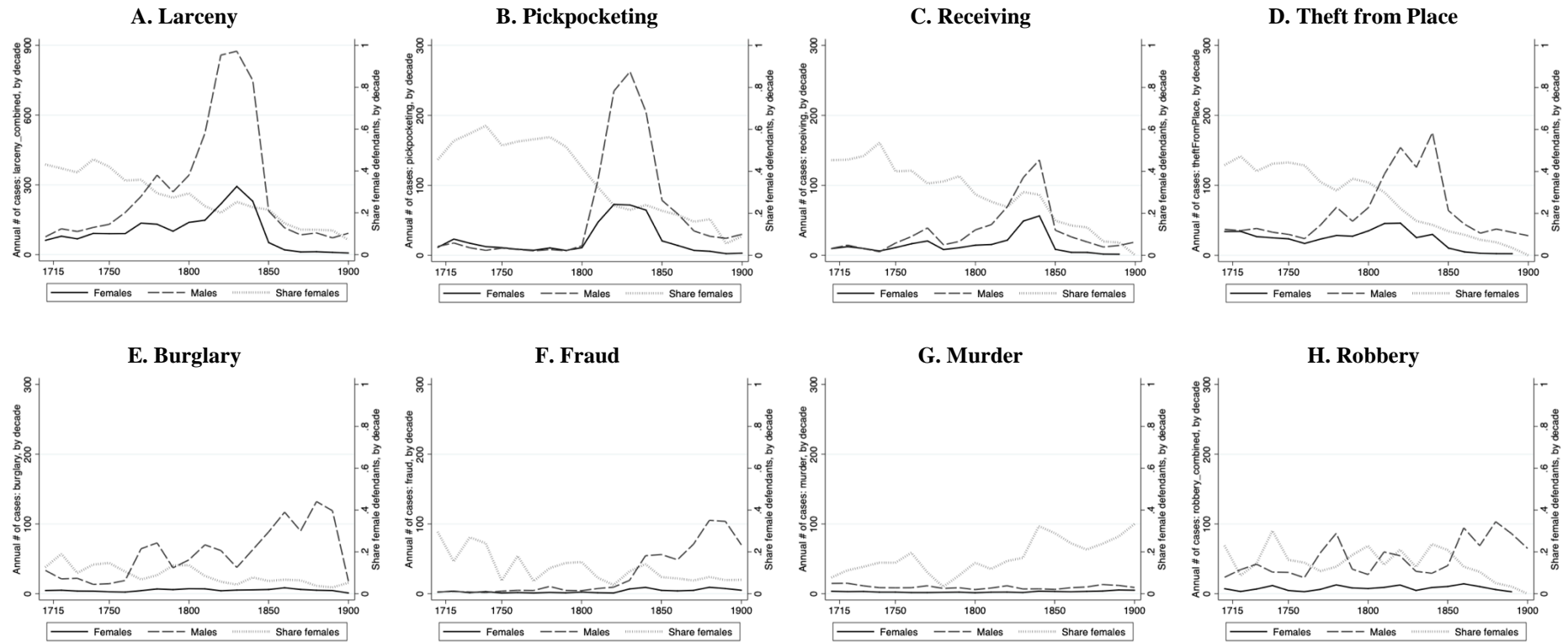


D. Adjusted gap in jury convictions, original offense



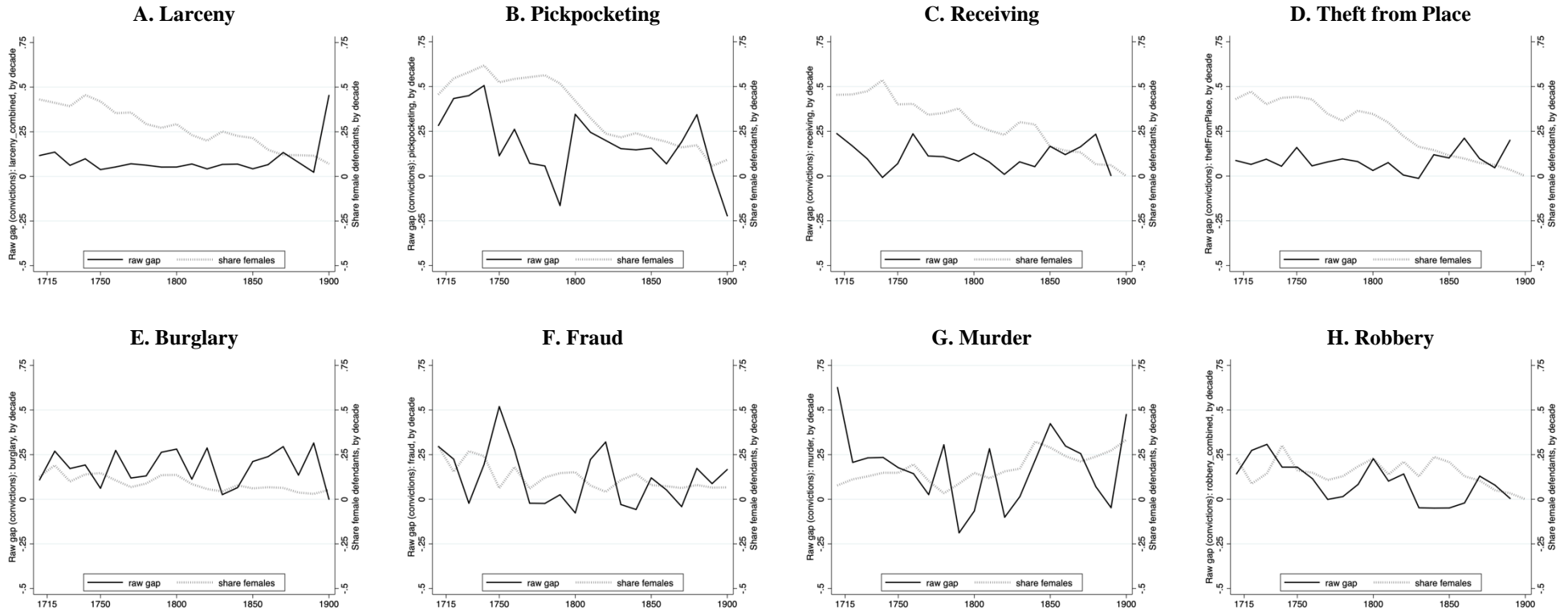
NOTE – Figure A (B) shows the average annual share of jury convictions of any (original) offense by gender (averaged by decade) as well as the raw gender gap relative to the male mean. Figures C and D plot the corresponding adjusted gender gap and 95% confidence interval as well as the raw gender gap. The adjusted gap corresponds to the estimated coefficients from equation (1) including controls for gender, number of defendants, capital eligibility and offense fixed effects. The sample is restricted to violent crimes only. SOURCE - *The Old Bailey Proceedings Online* and own calculations.

Appendix Figure 3. Number of Cases and Share of Female Defendants – Selected Offenses



NOTE – See Figure 1, Panels C and D. The figures show the corresponding figures for selected detailed offenses and averaged over decades instead of 5-year intervals. The solid and dashed black lines are the number of female and male cases, respectively. The dotted gray line is the share of female defendants.

Appendix Figure 4. Share of Female Defendants versus Raw Conviction Gaps by Offense Category

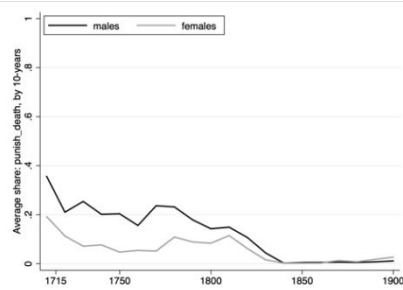


NOTE – In each figure, the black solid line (left axis) shows the raw gender gap in the average annual share of jury convictions of any offense (averaged by decade) and the grey dotted line the average annual share of female defendants across all trials for the specific offense (averaged by decade). Each figure corresponds to a specific offense category as indicated at the top of each one of them. SOURCE - *The Old Bailey Proceedings Online* and own calculations.

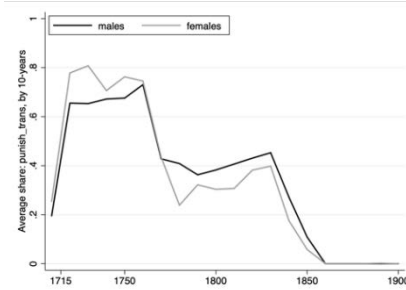
Appendix Figure 5. Raw and Adjusted Gender Gaps in Sentencing (All Crimes)

Panel A. Raw gaps

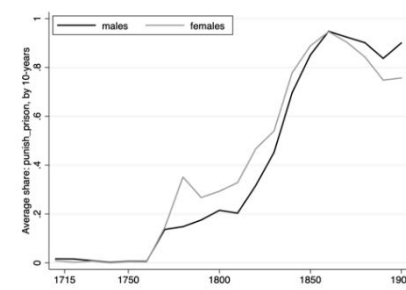
A1. Capital punishment



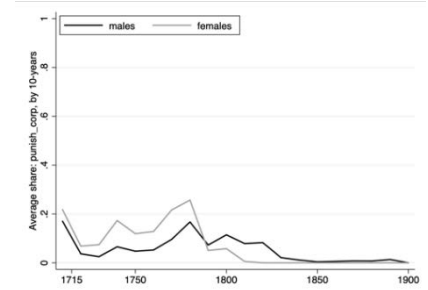
A2. Transportation



A3. Prison

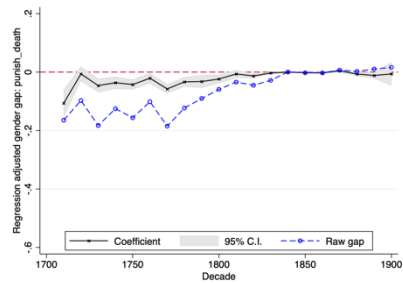


A4. Corporal punishment



Panel B. Adjusted gaps

B1. Capital punishment



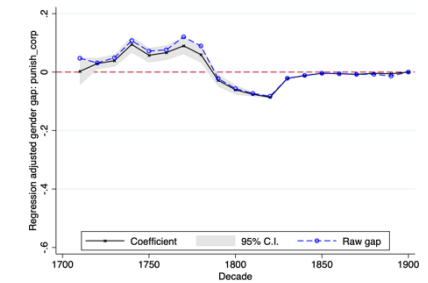
B2. Transportation



B3. Prison



B4. Corporal punishment



NOTE – Panel A shows the average share of cases for males (black line) and females (gray line) sentenced to capital punishment (A1), transportation (A2), prison (A3) and corporal punishment (A4), each by gender and for all offenses combined (averaged by decade and conditional on conviction). Panel B shows the adjusted gender gaps (solid black line) and 95% confidence intervals for four different sentencing types. The adjusted gap corresponds to the estimated coefficients from equation (1) including controls for gender, number of defendants, capital eligibility and offense fixed effects. The raw gender gap (blue, open circles, dashed line) is also shown. Figure B1 shows the results for capital punishment, B2 for transportation, B3 for prison and B4 for corporal punishment (by decade and conditional on conviction). Offense-specific reforms abolishing capital punishment started in 1808, transportation was step-wise abolished in 1853-57. SOURCE - *The Old Bailey Proceedings Online* and own calculations.

Appendix Table 1. Sample Creation

Detailed offence category (combined)	Drop	1715-1900		1765-1795		1803-1871		1845-1860	
		N	Share females	N	Share females	N	Share females	N	Share females
Against crown (major)	D1	36	0.22	3	1.00	28	0.00	10	0.00
Against crown (minor)	D1	134	0.05	68	0.00	65	0.00	12	0.17
Animal theft	D2	1281	0.02	579	0.02	2057	0.02	380	0.02
Arson		289	0.21	21	0.33	127	0.31	72	0.25
Assault		297	0.06	6	0.17	367	0.07	330	0.06
Bigamy		980	0.18	42	0.24	525	0.16	227	0.13
Burglary		4868	0.07	1781	0.08	3901	0.08	1366	0.07
Coining offences		3341	0.21	401	0.28	4432	0.30	3224	0.33
Embezzlement	D2	544	0.01	5	0.00	2625	0.02	1294	0.03
Forgery		2530	0.09	240	0.05	1806	0.08	1028	0.07
Fraud		3174	0.08	199	0.10	1369	0.11	997	0.09
Housebreaking		1088	0.16	238	0.20	2144	0.13	759	0.09
Infanticide (combined)	D2	471	0.88	25	0.96	252	0.93	140	0.95
Larceny (combined)		15605	0.33	11355	0.30	40278	0.24	5810	0.21
Libel		362	0.13	3	0.00	108	0.06	45	0.04
Mail	D2	742	0.03	9	0.00	263	0.01	203	0.02
Manslaughter		832	0.22	19	0.05	703	0.14	348	0.16
Missing	D3	157	0.21	27	0.07	77	0.18	18	0.11
Murder		1180	0.18	274	0.08	535	0.18	155	0.32
Other (combined)	D3	3818	0.36	324	0.24	1564	0.15	419	0.14
Perjury		601	0.23	133	0.17	272	0.19	179	0.17
Perverting justice		200	0.14	92	0.21	164	0.19	93	0.11
Pickpocketing		2362	0.38	430	0.56	10398	0.24	2284	0.22
Rape	D2	950	0.02	100	0.04	403	0.01	216	0.00
Receiving		1964	0.33	1132	0.35	4727	0.27	1433	0.26
Return from transportation		114	0.10	115	0.03	120	0.03	22	0.00
Riot		85	0.05	105	0.05	31	0.06	30	0.03
Robbery (combined)		4879	0.10	1928	0.12	3311	0.16	793	0.20
Sexual assault (combined)	D2	587	0.01	-	-	293	0.00	131	0.00
Shoplifting		1331	0.40	571	0.61	1500	0.45	110	0.27
Sodomy (combined)	D2	611	0.00	18	0.00	185	0.00	154	0.00
Stealing from master		1042	0.09	-	-	5792	0.25	3860	0.24
Theft from place		4659	0.35	2097	0.33	7477	0.22	1798	0.14
Wounding		2033	0.14	56	0.00	1494	0.18	1103	0.19

NOTE - The table shows the number of observations and share of female defendants by time period and offense categories. The last column indicates those offenses excluded from the analysis sample due to low observation numbers (D1), distinctly male or female offenses by definition or share of female defendants (D2), or because the nature of the offense is missing or not distinct (D3). Where applicable, we combine offense categories into one bigger category: against the crown – major (petty treason, piracy, seducing allegiance, treason), against the crown –minor (religious offenses, seditious libel, seditious words, tax offences), infanticide (concealing a birth, illegal abortion, infanticide), larceny (grand larceny, petty larceny, simple larceny), other (bankruptcy, barratry, conspiracy, extortion, keeping a brothel, kidnapping, other, threatening behavior, vagabonding), robbery (highway robbery, robbery), sexual assault (assault with intent, indecent assault), and sodomy (assault with sodomitical intent, sodomy). SOURCE - *The Old Bailey Proceedings Online* and own calculations.

Appendix Table 2. Gender Gaps in Convictions and Sentences in 50-Year Intervals

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
<i>Sample:</i>	<u>1715-1900</u>		<u><1751</u>		<u>1751-1800</u>		<u>1801-1850</u>		<u>≥1850</u>	
<i>Spec.:</i>	no ctrls.	obs. ctrls.	no ctrls.	obs. ctrls.	no ctrls.	obs. ctrls.	no ctrls.	obs. ctrls.	no ctrls.	obs. ctrls.
Panel A. Jury conviction of any charge										
Female	-0.0790***	-0.0803***	-0.1118***	-0.1234***	-0.0646***	-0.0803***	-0.0644***	-0.0688***	-0.0751***	-0.0944***
	(0.0176)	(0.0155)	(0.0312)	(0.0278)	(0.0142)	(0.0127)	(0.0199)	(0.0192)	(0.0240)	(0.0131)
Panel B. Jury conviction of original charge										
Female	-0.1170***	-0.0979***	-0.1415***	-0.1168***	-0.1144***	-0.1086***	-0.0718***	-0.0808***	-0.0669**	-0.0939***
	(0.0210)	(0.0145)	(0.0305)	(0.0157)	(0.0184)	(0.0084)	(0.0234)	(0.0217)	(0.0323)	(0.0178)
Panel C. Jury conviction of any charge or guilty by plea										
Female	-0.0950***	-0.0835***	-0.1139***	-0.1248***	-0.0653***	-0.0811***	-0.0690***	-0.0722***	-0.0880***	-0.0908***
	(0.0186)	(0.0157)	(0.0311)	(0.0274)	(0.0141)	(0.0123)	(0.0197)	(0.0187)	(0.0258)	(0.0136)
Panel D. Sentenced to harshest punishment available										
Female	-0.1440***	-0.0743***	-0.1008*	-0.0759***	-0.1436***	-0.1026***	-0.0959***	-0.0640**	-0.0630*	-0.0415**
	(0.0350)	(0.0160)	(0.0553)	(0.0215)	(0.0399)	(0.0120)	(0.0189)	(0.0257)	(0.0309)	(0.0158)
Observations										
Panels A-B	129,451	129,451	9,401	9,401	16,655	16,655	70,629	70,629	32,766	32,766
Panel C	192,050	192,050	15,838	15,838	30,556	30,556	95,318	95,318	50,338	50,338
Panel D	135,167	135,167	9,401	9,401	17,884	17,884	70,629	70,629	37,253	37,253

NOTE –The table shows the estimated coefficients when regressing the outcome (verdicts in panels A-C and sentences conditional on conviction in panels D to G) on a dummy for female defendants, corresponding to estimating equation (1). Columns (1) and (2) show the results for the whole sample period with/without observable controls (number defendants, capital eligibility, detailed offense category and year fixed effects); these are the baseline results shown for ease of comparison. Columns (3) and (4) show the respective results with/without observable controls for the sample period before 1751, columns (5) and (6) for 1751-1800, columns (7) and (8) for 1801-1850 and columns (9) and (10) for the sample period after 1850. Robust standard errors are clustered at the offense level and shown in parentheses. *** p<0.01, ** p<0.05, *p<0.1

Appendix Table 3. The Importance of *Feme Covert* in Explaining Gender Gaps

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Defendants:	Single	Multiple	Multiple	Multiple	Multiple	Two	Two
Gender:	All	All	Same	Mixed	Mixed	Mixed	Mixed
					case f.e.	case f.e. same name	case f.e. diff. name
Panel A. Jury conviction of any charge - all offenses							
Female	-0.0613*** (0.0173)	-0.1293*** (0.0143)	-0.1010*** (0.0184)	-0.1291*** (0.0234)	-0.1185*** (0.0323)	-0.1772*** (0.0298)	-0.0661 (0.0399)
Panel B. Jury conviction of original charge - all offenses							
Female	-0.0810*** (0.0150)	-0.1399*** (0.0140)	-0.1227*** (0.0193)	-0.1178*** (0.0171)	-0.1063*** (0.0226)	-0.1493*** (0.0362)	-0.0618** (0.0279)
Panel C. Jury conviction of any charge - property offenses							
Female	-0.0629** (0.0211)	-0.1418*** (0.0177)	-0.1101*** (0.0223)	-0.1477*** (0.0278)	-0.1429*** (0.0372)	-0.2023*** (0.0109)	-0.0842 (0.0489)
Panel D. Jury conviction of original charge - property offenses							
Female	-0.0785*** (0.0165)	-0.1445*** (0.0196)	-0.1217*** (0.0229)	-0.1284*** (0.0208)	-0.1225*** (0.0262)	-0.1633*** (0.0325)	-0.0782** (0.0323)
Observations							
Panels A-B	111,892	55,363	44,295	11,068	11,068	1,998	4,364
Panels C-D	86,671	41,091	33,335	7,756	7,756	1,469	3,134

NOTE – The table shows estimated coefficients when regressing the outcome (verdicts) on a dummy for female defendants, corresponding to estimating equation (1). Panel A and B show the results for all offenses and Panels C and D for property offenses, respectively. Column (1) is restricted to single defendants, (2) – (5) to multiple defendants and (6) – (7) to two defendant cases. Column (3) is restricted to same gender cases, (4) – (7) to mixed gender cases. Columns (5) – (7) include case fixed effects; (6) includes cases with defendants of the same surname, (7) of different surnames. All regressions include observable controls (see Table 3), except for when including case fixed effects. Robust standard errors are clustered at the offense level and shown in parentheses. *** p<0.01, ** p<0.05, *p<0.1

Appendix Table 4. Shift in Punishment Regimes (American Revolution) – Robust Standard Errors

Outcome:	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	any charge	<i>Jury conviction</i> original charge	lesser charge	death	<i>Judge sentence</i> transportation	prison	corporal
Panel A. All offenses							
Post 76-86	-0.0293 (0.0262)	-0.0001 (0.0263)	-0.0393** (0.0189)	0.0075 (0.0182)	-0.2367*** (0.0276)	0.1156*** (0.0140)	0.1360*** (0.0180)
Post 87-95	-0.0503* (0.0298)	-0.0151 (0.0296)	-0.0435** (0.0201)	-0.0042 (0.0211)	-0.4024*** (0.0325)	0.1107*** (0.0202)	0.0436*** (0.0161)
Female	-0.0899*** (0.0130)	-0.1190*** (0.0126)	0.0356*** (0.0096)	-0.0297*** (0.0080)	-0.0004 (0.0156)	-0.0242*** (0.0029)	0.0529*** (0.0122)
Post 76-86 x Fem.	0.0180 (0.0173)	0.0025 (0.0167)	0.0138 (0.0126)	-0.0297** (0.0118)	-0.1960*** (0.0184)	0.1358*** (0.0158)	0.0728*** (0.0183)
Post 87-95 x Fem.	0.0169 (0.0206)	0.0579*** (0.0199)	-0.0391*** (0.0132)	0.0248* (0.0141)	-0.0395 (0.0259)	0.1417*** (0.0172)	-0.1311*** (0.0157)
N	21,164	21,164	21,164	12,145	12,145	12,145	12,145
Panel B. Property offenses							
Post 76-86	-0.0134 (0.0278)	0.0216 (0.0279)	-0.0461** (0.0206)	0.0030 (0.0170)	-0.2624*** (0.0289)	0.1289*** (0.0143)	0.1551*** (0.0202)
Post 87-95	-0.0670** (0.0316)	-0.0221 (0.0315)	-0.0534** (0.0220)	0.0161 (0.0206)	-0.4183*** (0.0355)	0.1168*** (0.0207)	0.0433** (0.0179)
Female	-0.0874*** (0.0135)	-0.1077*** (0.0131)	0.0264*** (0.0101)	-0.0233*** (0.0076)	-0.0374** (0.0160)	-0.0109*** (0.0022)	0.0632*** (0.0129)
Post 76-86 x Fem.	0.0145 (0.0182)	-0.0072 (0.0177)	0.0211 (0.0137)	-0.0237** (0.0113)	-0.1470*** (0.0191)	0.1093*** (0.0170)	0.0610*** (0.0198)
Post 87-95 x Fem.	0.0089 (0.0217)	0.0433** (0.0210)	-0.0328** (0.0140)	0.0204 (0.0137)	-0.0137 (0.0278)	0.1448*** (0.0183)	-0.1500*** (0.0172)
N	17,575	17,575	17,575	10,313	10,313	10,313	10,313
Panel C. Violent offenses							
Post 76-86	0.0274 (0.1065)	0.0043 (0.0993)	0.0303 (0.0839)	0.0241 (0.1286)	-0.1098 (0.1208)	0.0453* (0.0237)	0.0051 (0.0152)
Post 87-95	-0.0076 (0.1338)	0.0165 (0.1271)	-0.0213 (0.0917)	0.0116 (0.1628)	-0.3329*** (0.1143)	0.1301 (0.0820)	-0.0003 (0.0021)
Female	-0.0173 (0.0649)	-0.1412** (0.0563)	0.1296** (0.0566)	-0.2509*** (0.0907)	0.2383*** (0.0779)	-0.0153 (0.0163)	0.0026 (0.0016)
Post 76-86 x Fem.	-0.0229 (0.0818)	0.0200 (0.0731)	-0.0503 (0.0669)	0.0153 (0.1116)	-0.1336 (0.0915)	0.0823* (0.0498)	0.0801** (0.0380)
Post 87-95 x Fem.	0.0138 (0.0893)	0.0928 (0.0813)	-0.0625 (0.0703)	0.1790 (0.1184)	-0.1678 (0.1056)	-0.0001 (0.0223)	0.0010 (0.0017)
N	2,281	2,281	2,281	1,138	1,138	1,138	1,138

NOTE – See Table 4. Robust standard errors are shown in parentheses. *** p<0.01, ** p<0.05, *p<0.1

Appendix Table 5. Abolition of Capital Punishment - Robust Standard Errors

Outcome:	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	any charge	<i>Jury conviction</i> original charge	lesser charge	death	<i>Judge sentence</i> transportation	prison	corporal
Panel A. All offenses							
Noncapital	0.0809*** (0.0066)	0.1693*** (0.0069)	-0.0932*** (0.0040)	-0.4138*** (0.0053)	0.1941*** (0.0074)	0.0927*** (0.0065)	0.0554*** (0.0025)
Female	-0.0649*** (0.0089)	-0.1064*** (0.0090)	0.0530*** (0.0074)	-0.1121*** (0.0102)	-0.0052 (0.0111)	0.0977*** (0.0100)	-0.0217*** (0.0024)
Noncapital x Female	-0.0032 (0.0096)	0.0320*** (0.0097)	-0.0465*** (0.0075)	0.1165*** (0.0102)	-0.0548*** (0.0116)	-0.0353*** (0.0107)	-0.0156*** (0.0026)
N	101,009	101,009	101,009	87,477	87,477	87,477	87,477
Panel B. Property offenses							
Noncapital	0.0204*** (0.0074)	0.1630*** (0.0079)	-0.1502*** (0.0048)	-0.3770*** (0.0059)	0.2035*** (0.0085)	0.0505*** (0.0074)	0.0493*** (0.0028)
Female	-0.0758*** (0.0101)	-0.1192*** (0.0106)	0.0563*** (0.0089)	-0.1110*** (0.0114)	-0.0117 (0.0130)	0.1080*** (0.0119)	-0.0281*** (0.0031)
Noncapital x Female	0.0017 (0.0108)	0.0390*** (0.0112)	-0.0500*** (0.0089)	0.1172*** (0.0114)	-0.0567*** (0.0136)	-0.0359*** (0.0126)	-0.0163*** (0.0033)
N	81,504	81,504	81,504	70,776	70,776	70,776	70,776
Panel C. Violent offenses							
Noncapital	0.1857*** (0.0292)	0.1222*** (0.0282)	0.0793*** (0.0196)	-0.8181*** (0.0317)	-0.0081 (0.0265)	0.6954*** (0.0361)	0.0216*** (0.0043)
Female	-0.1217*** (0.0248)	-0.1391*** (0.0218)	0.0264 (0.0196)	-0.1471*** (0.0373)	0.0974*** (0.0337)	0.0239 (0.0263)	0.0019 (0.0044)
Noncapital x Female	0.1158*** (0.0293)	0.1150*** (0.0274)	-0.0091 (0.0233)	0.1321*** (0.0374)	-0.1165*** (0.0356)	0.0174 (0.0301)	-0.0156*** (0.0048)
N	8,702	8,702	8,702	5,578	5,578	5,578	5,578

NOTE – See Table 5. Robust standard errors are shown in parentheses. *** p<0.01, ** p<0.05, *p<0.1

Appendix Table 6. Abolition of Transportation - Robust Standard Errors

Outcome:	(1) <i>Jury conviction</i> any charge	(2) <i>Jury conviction</i> original charge	lesser charge	(3) death	(4) <i>Judge sentence</i> transportation	(5) <i>Judge sentence</i> prison	(6) corporal
Panel A. All offenses							
Post 1853	-0.0089 (0.0208)	-0.0024 (0.0215)	-0.0065 (0.0093)	0.0028 (0.0024)	-0.2689*** (0.0110)	0.2463*** (0.0137)	-0.0022* (0.0012)
Female	-0.0529*** (0.0099)	-0.0541*** (0.0100)	0.0012 (0.0035)	-0.0001 (0.0010)	-0.0705*** (0.0077)	0.0473*** (0.0087)	-0.0100*** (0.0010)
Post 1853 x Female	-0.0307* (0.0171)	-0.0335* (0.0173)	0.0029 (0.0068)	-0.0022 (0.0018)	0.0781*** (0.0090)	-0.0782*** (0.0119)	0.0098*** (0.0010)
N	18,591	18,591	18,591	19,875	19,875	19,875	19,875
Panel B. Property offenses							
Post 1853	-0.0322 (0.0294)	-0.0348 (0.0303)	0.0025 (0.0107)	-0.0002 (0.0002)	-0.3022*** (0.0130)	0.2832*** (0.0165)	-0.0049*** (0.0015)
Female	-0.0571*** (0.0118)	-0.0609*** (0.0118)	0.0038 (0.0025)	-0.0002 (0.0005)	-0.0702*** (0.0090)	0.0473*** (0.0101)	-0.0128*** (0.0013)
Post 1853 x Female	-0.0398 (0.0257)	-0.0482* (0.0260)	0.0085 (0.0079)	0.0019 (0.0017)	0.0707*** (0.0113)	-0.0867*** (0.0172)	0.0112*** (0.0013)
N	11,866	11,866	11,866	13,380	13,380	13,380	13,380
Panel C. Violent offenses							
Post 1853	0.0846 (0.0555)	0.1744*** (0.0574)	-0.0898** (0.0433)	0.0148 (0.0217)	-0.2264*** (0.0368)	0.1873*** (0.0533)	-0.0007 (0.0007)
Female	-0.0517* (0.0299)	-0.0432 (0.0303)	-0.0085 (0.0276)	-0.0001 (0.0116)	-0.0611* (0.0314)	0.0454 (0.0351)	-0.0011 (0.0011)
Post 1853 x Female	-0.0087 (0.0462)	-0.0145 (0.0475)	0.0058 (0.0383)	-0.0233 (0.0166)	-0.0285 (0.0376)	0.0470 (0.0469)	0.0007 (0.0007)
N	2,546	2,546	2,546	1,887	1,887	1,887	1,887

NOTE – See Table 6. Robust standard errors are shown in parentheses. *** p<0.01, ** p<0.05, *p<0.1

Appendix Table 7. Shift in Punishment Regimes (American Revolution) - Robustness to Trends

Outcome:	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	any charge	<i>Jury conviction</i> original charge	lesser charge	death	<i>Judge sentence</i> transportation	prison	corporal
Panel A. No year fixed effects and no trend (all offenses)							
Halt (1776-1786)	-0.0015 (0.0116)	0.0305** (0.0123)	-0.0311*** (0.0089)	0.0231 (0.0178)	-0.3998*** (0.0771)	0.2279*** (0.0387)	0.1045** (0.0380)
Australia (1787-1795)	-0.0205 (0.0120)	0.0233 (0.0228)	-0.0455* (0.0264)	-0.0171 (0.0134)	-0.1715*** (0.0585)	0.0872*** (0.0197)	0.0326** (0.0143)
Female	-0.0894*** (0.0164)	-0.1187*** (0.0102)	0.0361** (0.0130)	-0.0304 (0.0279)	0.0013 (0.0628)	-0.0234* (0.0127)	0.0513*** (0.0174)
Halt x Female	0.0136 (0.0161)	-0.0074 (0.0140)	0.0198 (0.0164)	-0.0237 (0.0183)	-0.2502*** (0.0515)	0.1716*** (0.0351)	0.0684*** (0.0165)
Australia x Female	0.0182 (0.0179)	0.0586*** (0.0132)	-0.0384** (0.0139)	0.0251 (0.0240)	-0.0586 (0.0351)	0.1435*** (0.0385)	-0.1359*** (0.0459)
N	21,164	21,164	21,164	12,145	12,145	12,145	12,145
Panel B. Linear trend (all offenses)							
Halt (1776-1786)	0.0098 (0.0095)	0.0108 (0.0090)	0.0054 (0.0121)	0.0423 (0.0345)	-0.5152*** (0.0785)	0.3265*** (0.0410)	0.0746* (0.0376)
Australia (1787-1795)	-0.0006 (0.0166)	-0.0114 (0.0141)	0.0190 (0.0213)	0.0157 (0.0315)	-0.3690*** (0.0564)	0.2559*** (0.0239)	-0.0186 (0.0191)
Female	-0.0896*** (0.0166)	-0.1185*** (0.0102)	0.0357** (0.0130)	-0.0305 (0.0282)	0.0023 (0.0641)	-0.0242* (0.0136)	0.0516*** (0.0171)
Halt x Female	0.0130 (0.0157)	-0.0063 (0.0142)	0.0178 (0.0169)	-0.0248 (0.0189)	-0.2437*** (0.0539)	0.1660*** (0.0370)	0.0701*** (0.0157)
Australia x Female	0.0187 (0.0182)	0.0577*** (0.0132)	-0.0367** (0.0140)	0.0260 (0.0246)	-0.0640* (0.0371)	0.1482*** (0.0379)	-0.1373*** (0.0456)
N	21,164	21,164	21,164	12,145	12,145	12,145	12,145
Panel C. Offense-specific linear trend (all offenses)							
Halt (1776-1786)	0.0078 (0.0100)	0.0132 (0.0089)	0.0007 (0.0128)	0.0449 (0.0360)	-0.5287*** (0.0739)	0.3320*** (0.0386)	0.0782** (0.0372)
Australia (1787-1795)	-0.0089 (0.0159)	-0.0117 (0.0141)	0.0103 (0.0184)	0.0237 (0.0345)	-0.3917*** (0.0477)	0.2644*** (0.0231)	-0.0143 (0.0195)
Female	-0.0950*** (0.0193)	-0.1144*** (0.0098)	0.0252* (0.0130)	-0.0272 (0.0275)	-0.0179 (0.0552)	-0.0151* (0.0079)	0.0573*** (0.0149)
Halt x Female	0.0202 (0.0175)	-0.0099 (0.0142)	0.0297 (0.0184)	-0.0295 (0.0207)	-0.2177*** (0.0416)	0.1542*** (0.0306)	0.0637*** (0.0132)
Australia x Female	0.0300 (0.0232)	0.0526*** (0.0123)	-0.0192 (0.0196)	0.0187 (0.0237)	-0.0293 (0.0219)	0.1328*** (0.0430)	-0.1468*** (0.0425)
N	21,164	21,164	21,164	12,145	12,145	12,145	12,145

NOTE – The table shows the robustness of the results for all offenses shown in Panel A of Table 4. See the notes of Table 4 for details regarding the sample period and baseline specifications. Panel A omits year fixed effects, Panel B includes a linear trend (no year fixed effects) and Panel C includes offense-specific linear trends (no year fixed effects). Robust standard errors are clustered at the offense level and shown in parentheses. *** p<0.01, ** p<0.05, *p<0.1

Appendix Table 8. Abolition of Transportation - Robustness to Trends

Outcome:	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	any charge	<i>Jury conviction</i> original charge	lesser charge	death	<i>Judge sentence</i> transportation	prison	corporal
Panel A. No year fixed effects and no trend (all offenses)							
Post 1853	-0.0151 (0.0153)	-0.0085 (0.0187)	-0.0066 (0.0074)	0.0018 (0.0015)	-0.2144*** (0.0427)	0.2054*** (0.0442)	-0.0078*** (0.0023)
Female	-0.0520** (0.0239)	-0.0535** (0.0246)	0.0014 (0.0048)	-0.0000 (0.0013)	-0.0702*** (0.0171)	0.0473** (0.0212)	-0.0100*** (0.0020)
Post 1853 x Female	-0.0302 (0.0210)	-0.0336 (0.0249)	0.0034 (0.0089)	-0.0023 (0.0028)	0.0785** (0.0281)	-0.0786** (0.0288)	0.0097*** (0.0016)
N	18,591	18,591	18,591	19,875	19,875	19,875	19,875
Panel B. Linear trend (all offenses)							
Post 1853	0.0063 (0.0188)	0.0086 (0.0215)	-0.0023 (0.0054)	0.0032 (0.0030)	-0.1221*** (0.0385)	0.1130*** (0.0385)	-0.0120*** (0.0033)
Female	-0.0524** (0.0238)	-0.0537** (0.0245)	0.0014 (0.0048)	-0.0000 (0.0013)	-0.0702*** (0.0170)	0.0472** (0.0212)	-0.0100*** (0.0020)
Post 1853 x Female	-0.0302 (0.0211)	-0.0336 (0.0250)	0.0034 (0.0089)	-0.0024 (0.0028)	0.0774** (0.0277)	-0.0775** (0.0280)	0.0098*** (0.0017)
N	18,591	18,591	18,591	19,875	19,875	19,875	19,875
Panel C. Offense-specific linear trend (all offenses)							
Post 1853	0.0016 (0.0183)	0.0028 (0.0210)	-0.0012 (0.0060)	0.0030 (0.0030)	-0.1116*** (0.0359)	0.1016*** (0.0357)	-0.0124*** (0.0033)
Female	-0.0524** (0.0236)	-0.0542** (0.0242)	0.0018 (0.0048)	0.0001 (0.0013)	-0.0587*** (0.0174)	0.0355 (0.0221)	-0.0100*** (0.0020)
Post 1853 x Female	-0.0304 (0.0204)	-0.0325 (0.0235)	0.0022 (0.0079)	-0.0024 (0.0029)	0.0445** (0.0184)	-0.0437* (0.0223)	0.0099*** (0.0018)
N	18,591	18,591	18,591	19,875	19,875	19,875	19,875

NOTE – The table shows the robustness of the results for all offenses shown in Panel A of Table 6. See the notes of Table 6 for details regarding the sample period and baseline specifications. Panel A omits year fixed effects, Panel B includes a linear trend (no year fixed effects) and Panel C includes offense-specific linear trends (no year fixed effects). Robust standard errors are clustered at the offense level and shown in parentheses. *** p<0.01, ** p<0.05, *p<0.1