

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 23, 2017

PRELIMINARY AND INCOMPLETE DRAFT

Abstract: This paper studies the gender gap in defendant plea behavior, jury verdicts and judge sentences for more than 200,000 trials spanning 200 years at the Old Bailey Central Criminal Court of London. We find a gender gap favoring females at every stage of the judicial system that persists throughout the sample period. Females were less likely to be found guilty (9.4% - 20.3%) and, conditional on guilt, less likely to receive the harshest punishment available at the time (i.e. death penalty, transportation, or prison depending on the offense and year). Oaxaca decompositions suggest that gender differences in observable case characteristics do not explain any of the conviction gap and little of the gap in sentencing. We argue and provide empirical evidence that unobservable differences in case characteristics, including females as better defendants and non-random assignment of females to more lenient judges and juries, do not explain the remaining gaps. The remainder of the paper assesses whether the gender gap is driven by differential attitudes towards females: our results suggest that statistical discrimination and a preference to protect ‘mothers’ can likely be ruled out while the girlfriend theory and preference-based discrimination may explain at least part of the gap.

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. 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 sharp contrast to the well-known gender gaps favoring men in the labor market, numerous studies today document a criminal justice system that treats females more leniently than males.¹ This has been especially seen with respect to sentencing. 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. 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. This gender gap continues to expand, as the average male sentence length is between 0.9 and 1.6 times that for females; the sentence length ratio is less than one for just one of ten crime categories (criminal damage).²

The current paper contributes to the understanding of this modern day criminal justice gender gap by studying its evolution over a dynamic 200-year period. In particular, we study the raw and adjusted gender gaps in defendant pleas, jury convictions, and judge sentences from 1715 to 1913 at the Old Bailey Central Criminal Court in London. Our contributions are three-fold: we look at (i) the dynamics of the gender gap for (ii) various stages of the justice system and (iii) provide evidence for which channels can and cannot contribute to the gap.

More specifically, this is the first paper to study the dynamics of the gender gap over an extended period (let alone 200 years). Most existing research focuses on modern-day static snapshots; a number of studies use U.S. federal court data for a three to ten year period between 1990 and 2010 (Starr, 2015; Mustard, 2001; Schanzenbach, 2005; Sorensen, et al, 2012). Yet, similar findings are also seen in state courts: Butcher et al. (forthcoming) find that female felons in Kansas from 1998 to 2011 receive lighter sentences than men.^{3 4} Second, we can study the

¹ See Blau and Kahn (forthcoming) for a recent review of the U.S. gender wage gap. They highlight that though the gender wage gap has declined in recent decades, there is still a significant gap, which is difficult to explain. Almost none of the wage gap today is explained by differences in human capital. While occupation and industry differences explains some of it, a significant portion is unexplained.

² The gender gaps in the rate of immediate custodial sentences by offense category are: violence against persons (2.1), sex (1.1), burglary (1.7), robbery (1.6), theft (1.6), fraud and forgery (2.3), criminal damage (1.5), drugs (1.1), motoring (3.2), and other (1.5). Sentencing statistics are calculated from the data underlying a Ministry of Justice publication, Sentencing Statistics England and Wales 2009. See Tables 2i and 2j of the supplementary tables <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 Similar periods are studied in other jurisdictions, including 2000-2003 sentences in France (Philippe, 2017).

⁴ Given data challenges, it is unsurprising that there are not many papers studying criminal justice gender gaps in a historical context; one exception is Bodenhorn (2009), who finds that females received shorter sentences in 19th century Pennsylvania. Females were not the focus of the analysis, however, and comprised just 4% of the sample of 10,000 individuals sentenced to two state prisons in the mid-1800s. Bodenhorn (2009) also considers whether age, occupation, and race affects sentencing. Another historical paper that considers the importance of socioeconomic status is Vickers (2016), who looks at whether social status affects sentencing for (only) males in the Assize or Quarter Sessions in England and Wales in 1870, 1883, and 1910.

gender gap in various stages of the criminal justice process – plea, verdict, and sentence – for the same sample of defendants.⁵ With the notable exception of Starr (2015), previous research generally focused on a single stage (typically sentencing). Sorensen et al. (2012) find an unexplained gender gap of 30% in U.S. federal sentence lengths while Mustard (2001) and Schanzenbach (2005) find gaps of 12% and 11%, respectively.⁶ Starr (2015) estimates much larger gender gaps in sentence length (63%); she attributes the larger magnitude to the fact that she accounts for small (around 2%) but significant gender gaps at the charge and conviction stages. Other studies support the idea that gender gaps exist at earlier judicial stages. For instance, Stolzenberg and D'Alessio (2004) find females to be at least 9% less likely to be arrested (depending on offense) and Spohn and Spears (1997) and Shermer and Johnson (2010) both find a gender gap in plea bargaining.

Much of the modern day criminal justice gender gap is unexplained by observable differences in case characteristics.⁷ Starr (2015) puts forth a number of potential explanations, including (i) unobservable case differences, (ii) the ‘girlfriend’ theory, where female defendants in multiple defendant cases are not deemed responsible, (iii) concern for a female defendant’s maternal responsibilities, (iv) statistical discrimination, and (v) preference based discrimination (in the form of chivalry or paternalism). She argues that unobservables are not the driving factor in the modern-day context given the detailed nature of the observable data, and that while the girlfriend theory and maternal responsibilities may explain some of the gap, they explain far from all of it. Discrimination is left as a potential, but untested, explanation.⁸ Thus, our final contribution is to provide evidence on which channels may or may not explain the gender gap, including empirical tests with regards to the role of statistical discrimination.

Our data come from *The Proceedings of the Old Bailey* from 1715 to 1913, which were published (quasi-officially since the late 1700s) after each monthly court session of the Central Criminal Court in London. This unique data source has since been digitized by *The Old Bailey*

⁵ While we can look at multiple stages of court decisions, we cannot estimate the gender gap for the first stage grand jury decisions that determine which cases go to trial nor for the last stage judge pardons.

⁶ The estimated modern-day gap in French sentence lengths is 33% (Philippe, 2017).

⁷ For instance, 30% of the gap estimated by Butcher et al (forthcoming) is unexplained by observable differences in characteristics.

⁸ Other studies consider how the sentencing gender gap is affected by the gender composition of the judges. Schanzenbach (2005) and Philippe (2017) find that increasing the proportion of female judges decreases the gender gap. All judges and jurors during our sample period, however, are male. Though not focused on the gender gap, Anwar et al (2016) study the impact of adding females to the London jury pool in 1921. They find that this reform impacts the conviction rate for what can be considered more ‘female salient’ cases: there is (i) an increase in convictions for sex offenses, (ii) a widening of the differential conviction rate for male and female victim cases, and (iii) a reduction in the chance of conviction for females charged with other (including abortion) offenses. However, these female salient (sex and abortion) offenses are excluded from the current paper.

Proceedings Online, and includes (tagged) information identifying the case, session date, defendant's name and gender, and detailed offense, verdict, and sentencing categories. In addition, we manually coded judge, jury and juror names from 1750 to 1822 and defendant criminal history from the 1830s onwards (that is, for the periods when these variables are available). Our analysis sample consists of 202,569 trials across 23 (non-distinctly male or female) offense categories for this approximately 200-year period.

Our empirical analysis begins by documenting the raw (unadjusted) gender gap in jury verdicts, defendant pleas, and judge sentences. Throughout this 200-year period, females are persistently less likely to be convicted by the jury: specifically, they are 12 percentage points (20.3%) less likely to be convicted prior to 1751, eight percentage points (13.5%) from 1751 to 1800, seven percentage points (9.4%) from 1801 to 1850 and nine percentage points (14.1%) after 1850. In addition, males are persistently more likely (20-30%) to plead guilty throughout the 19th century (once pleading is part of the criminal justice system). To study the gender gap in sentencing, it is essential to consider the dramatically changing penal system during this period. In the early 1700s, capital punishment and transportation to the Americas dominated, while prison was virtually non-existent. Reforms in the 1800s abolished capital punishment offense by offense, such that transportation (to Australia) became the harshest available punishment. Prison was increasingly used over time, until it was the harshest punishment upon the abolition of transportation in the 1850s. Much of our empirical analysis of sentencing, thus, defines the outcome as being sentenced to the harshest punishment available in the current sentencing regime. The adjusted gender gaps in sentencing are such that females were, on average, 7.5 percentage points less likely to be sentenced to this harshest punishment.

Regression-adjusted gaps and an Oaxaca decomposition highlight that observable differences in case characteristics (including detailed offense category, number of offenses, capital offense status, and criminal history and defendant age) explain little of the raw gender gap. Specifically, observable characteristics explain 0% of the raw gap in conviction; while up to one-third of the gap in being sentenced to the harshest punishment available is explained by observables, these estimates are not statistically significant.

The remainder of the paper is dedicated to assessing whether this unexplained share of the gender gap is due to unobservable case differences for male and female defendants or differential attitudes towards female defendants. These differential attitudes could take the form of statistical or taste based discrimination, or arise out of the idea that mothers are needed at home to take care of the children, or that females (particularly those part of a married couple) are not capable of a crime themselves (the so-called girlfriend theory).

Though conditioning on observable characteristics had little impact on the estimated gender gap, there may still be unobservable differences in case characteristics across gender. Rather than relying on the typical argument that little selection on observables implies little selection on unobservables, we empirically test for the potential importance of the latter using as a proxy the number of words per trial printed in the Proceedings. Since testimony is often published verbatim, it is credible that word count captures characteristics observable to the jury but unobservable to us – like testimony length, witness credibility, or more information about the severity of the offense. Controlling for word count over and above observable controls in fact has little to no effect on the estimated gender gap. The robustness of the results to controlling for word count also suggests that one should not be concerned about unobserved defendant quality. That is, word count should at least partially capture the possibility that females make ‘better’ defendants, perhaps by providing more convincing testimony or more witnesses. This channel is also unlikely given the extremely short time allotted to trials and general difficulties in enforcing the attendance of witnesses, making it hard to imagine how one would have distinguished themselves as a ‘good’ defendant. Finally, using judge and jury fixed effects for a sub-sample of about 70 years, we also rule out that these gender gaps are driven by an unobserved systematic assignment of female defendants to more lenient judges and juries.

In terms of attitudinal differences towards female defendants, we argue that stories of statistical discrimination and a preferential treatment of female defendants because they are ‘mothers’ with responsibilities for childcare are both unlikely explanations of the remaining gender gaps. With respect to statistical discrimination, we rely on the fact that the share of female defendants declined from almost 40% to almost 10% over our sample period. If this decline is due to a true decrease in female criminality, then one would expect a widening of the gap over time. If this downward trend of female defendants is due to a shift of the least serious criminals to alternative courts, where female defendants are over-represented, then one would expect a narrowing of the gap over time; in this case, the remaining female defendants may be perceived as more criminal. The gender gap, however, does not trend one way or another over time, and in fact, remains relatively stable. This is even the case for the few offenses for which there is an opposite, increasing trend over time. In a similar spirit, we look across offenses that vary in their share of female defendants, but do not find consistent evidence of a larger gender gap for more male-dominated offenses. We rule out the ‘preferential treatment of mother’ channel primarily by using the Old Bailey Corpus Online, a search engine of all spoken text in the Proceedings and assessing the extent to which testimony spoke of ‘children’. This word was spoken in just over 1000 trials (a small fraction of all trials), and not substantially over-

represented by female defendants.

We empirically test the possibility that females are treated leniently because they are not deemed to be the responsible party (the girlfriend theory) by looking at subsamples of single defendant cases, and multiple defendant cases with same and mixed gender composition. We conclude that the girlfriend theory cannot be the predominant explanation overall, as a large and significant gap in verdicts exists in single defendant cases. However, for multiple defendant cases, we argue that this channel explains some, though not all, of the gender gap which is larger in mixed than for same gender cases for a number of offenses. The gender gap also tends to be larger for defendant pairs with the same name (a signal of spouses).

Finally, we consider preference based discrimination, which may take the form of paternalism or a desire by all male judges and jurors to ‘protect the weaker sex’. This seems a reasonable, if not likely, channel given the differing and persistent perceptions of the natures of men (physically and mentally strong, intelligent, determined) and women (passive, emotional, inferior and in need of care) and their differing societal roles. Did this male responsibility to care for the ‘weaker’ sex carry over to the male jurors and judges in the courtroom? While such preference-based discrimination implies that *convicted* females have higher quality of evidence against them than males, we cannot (yet) empirically test this given the difficulties in measuring quality of evidence.⁹

In summary, we find large and significant gender gaps that persist across the 200-year period and are, for the most part, unexplained by observable differences in case characteristics. We rule out three channels – unobservable case differences, preferential treatment of mothers, and statistical discrimination – as potential explanations underlying these gender gaps. Two channels – the girlfriend theory and preference-based discrimination – cannot be ruled out as explaining (at least part of) it.

What implications does our study have for the understanding of the criminal justice gender gap today? A first conclusion is that the gender gap is not a modern day phenomenon, but rather dates back more than 300 years. A second conclusion is that female representation in the courtroom as judges, jurors, and attorneys did not eliminate the gender gap. In terms of the mechanisms, we are left with the idea that the gender gap is largely a result of attitudes towards women as the weaker sex and of paternalism. Though attitudes towards females have clearly changed over the last 300 years, this study, combined with the current body of research, raises the question of how far we have really come.

⁹ We are currently exploring the possibility of additional data collection to allow for empirical tests of this channel.

The remainder of the paper proceeds as follows. Section 2 discusses the English criminal justice system and the role and perceptions of females throughout the 18th and 19th centuries. Section 3 describes the data and graphically depicts the raw gender gap in plea, conviction and sentencing. Section 4 presents the regression adjusted gender gaps and demonstrates that they are not driven by observable differences in characteristics. Section 5 considers the importance of unobservable differences while Section 6 considers differential attitudes and discrimination as potential explanations. Section 7 concludes.

2. Historical Context and Institutional Background

2.1. The English Criminal Justice System in the 18th and 19th Centuries

This study traces out the gender gap in jury trials at the Old Bailey Criminal Court between 1715 and 1900. The Old Bailey trialed the most serious crimes, including all felonies, for London and the surrounding county of Middlesex; the catchment area increased over time, especially in the 1830s with the addition of Essex. Before being sent to a jury trial at the Old Bailey, a Grand Jury decided whether there was sufficient evidence to proceed. This Grand Jury decision unfortunately occurs outside the scope of our data. That is, our analysis of gender gaps is conditional on a case having reached ‘trial’ at the Old Bailey. To the extent that there is a gender gap at earlier stages, including crime reporting or Grand Jury decisions, our analysis will underestimate the overall criminal justice gender gap. This is also a concern, perhaps to even a greater extent, for most modern day studies that are only focused on the sentencing stage.

The Old Bailey jury trials were organized in sessions lasting at least a few days, during which time a jury decided many consecutive cases.¹⁰ Potential jurors were listed on annually updated master lists. To be eligible, one had to be male, aged 21 to 60, reside in England, and meet income/wealth qualifications.¹¹ From the pool of jurors summoned to the court (from the master list), 12 were randomly drawn to sit on a jury. Separate juries were seated for London and Middlesex crimes.¹² After listening to testimony, the seated jury had to reach a unanimous verdict. The most common outcomes were acquittal, guilty, or guilty of a lesser offense. The judge decided the sentence (though the jury could recommend mercy). Until 1840, all sentences were announced on the last day of the session rather than immediately after each trial.

¹⁰ See Bindler and Hjalmarsson’s (2017b) analysis of the path dependency of these consecutive jury decisions.

¹¹ There were some changes in the upper age limit and level of wealth qualifications during this period. Women were excluded until the Sex Disqualification (Removal) Act of 1919.

¹² For more details, see e.g. Beatie (1986) for the Jury Act of 1730 and Bentley (1998) for the Juries Act of 1825.

There were numerous reforms of the English penal system between 1715 and 1900.¹³ The 18th century reforms primarily increased the number of capital offenses; in the early 1800s, more than 200 offenses (including sub-categories) were capital. These reforms were largely meant to protect the property of the wealthy classes, which is reflected in the fact that even today's petty crimes of pickpocketing or shoplifting were capital. In contrast, reforms in the first half of the 19th century abolished the death penalty for most offenses, sentencing convicts to penal transportation instead. Transportation to the Americas began in 1718 but was unexpectedly halted during the American Revolution; it was reinstated (to Australia) in 1786. Transportation – particularly to Australia – was perceived as a harsh punishment, involving a long and dangerous voyage, and hard labor and harsh discipline upon arrival. Viewed increasingly as inhumane, transportation was abolished in the 1850s. The idea of a prison sentence emerged when it was used as a substitute to transportation during the Revolution, when prisoners were held in the ships' hulks and put to work dredging the River Thames. During the latter half of the 1800s, prisons were built around London. Finally, convicts could be sentenced to corporal punishment such as the pillory or whipping, or other, miscellaneous punishments, e.g. fines, which could be combined with other punishments. Bindler and Hjalmarsson (2017a) provide additional details about this historical period, and suggests the existence of a gender gap: abolishing capital punishment differentially affected jury verdicts by defendant gender.

As is perhaps clear from the above description, the only role played by women during the 1700s and 1800s in the criminal justice system was defendant. All decision makers – jurors, judges and attorneys – were male. This began to change with the Sex Disqualification (Removal) Act of 1919.

2.2. The Role of Women in Pre- and Post-Industrial London

Criminal justice was not the only changing institution in 18th and 19th century London. Marked by the Industrial Revolution, this period also saw dramatic population growth, urbanization, and immigration. London's population increased from about 750,000 in 1760 to over one million in 1801 (the year of the first census) to seven million in 1911. In 1841, more than one-third of Londoners were not born there, and had often migrated there for work.¹⁴

Detailed statistics characterizing the labor market through these two centuries, especially for females, are particularly hard to come by. The first census was collected in 1801, but

¹³ Some – the 1836 entitlement to defense attorneys for felonies and the 1827 shift of the burden of proof to the prosecution, introducing the presumption of innocence – were aimed at improving the defendant's rights.

¹⁴ See <https://www.oldbaileyonline.org/static/Population-history-of-london.jsp>, viewed on August 30, 2017.

occupations not recorded until 1841. Even then, however, householders were advised that the professions of wives need not be indicated (Horrell and Humphries, 1995). After 1841, we use decennial censuses to paint a ‘coarse’ picture of post-industrial female labor market participation. For the pre-industrial period, we rely on anecdotal sources.

In the early to mid-1700s, women in London had few occupational choices. According to the Old Bailey Online, about half of women and 5% of men were in domestic service. Other jobs common to females included needlework, laundry, and street selling.¹⁵ Nicholas and Oxley (1993) characterize the occupations of women convicts from 1817 to 1840. The ten largest occupations included: housemaid (28.4%), maids of ‘all work’ (25.4%), kitchenmaid (11%), nursemaid (9.7%), cook (8.7%), laundress (8.3%), dairymaid (7.5%), needlewoman (6.9%), country servant (5.6%) and laundry maid (3.9%). Domestic service still dominated the female labor market at this time, at least for the type of ‘criminal’ individuals in our sample.

Finally, we use data for the London and Middlesex population from the 1851, 1861, 1881, 1891, 1901 and 1911 censuses to look at female labor market participation by age.¹⁶ Panel A of Appendix Figure 1 shows the shares of men and women younger than 18, between 18 and 34, and older than 34 who participate in the labor market. There is no labor force participation gender gap for juveniles. However, a large participation gap emerges for the 18-34 group that remains fairly persistent over the period: 47% of females versus 94% of men in 1851. This gap expands for the older than 34 subsample, such that men are 57 percentage points more likely to participate in the labor market. The increasing gap is driven by lower female participation, which could be due to the woman as responsible for the home and children and male as ‘breadwinner’ attitude of the time (Horrell and Humphries, 1995).

Panel B of Appendix Figure 1 shows the share of females by broad sector. The highest share is seen amongst service workers (almost 80%) while the lowest is for clerical and related occupations. Appendix Table 1 lists the 20 most common occupations, representing more than 50% of the labor force, in 1851 and 1901, respectively, and the share of females in each. A few phenomena are apparent. First, there is substantial overlap in occupations across this 50-year period. Only three occupations from 1851 are not amongst the most common in 1901: weavers, milliners (hat makers) and farmworkers are replaced by pipe fitters, freight handlers, and teachers. Second, regardless of period, men and women are segregated by occupation. Finally, as in pre-industrial times, women are predominantly employed in domestic service: the largest

¹⁵ <https://www.oldbaileyonline.org/static/London-life18th.jsp#socialandoccupationalstructure>, August 30, 2017.

¹⁶ The data was obtained from the North Atlantic Population Project, <https://www.nappdata.org/napp/>.

(by far) single category for women in both periods is house servants and maids.

2.3. Persistent Perceptions of Females as the Weaker Sex

Societal attitudes towards females as the ‘weaker sex’ remain remarkably persistent during this 200-year period. Well-known (male and female) 18th and 19th century authors make clear – sometimes in a call for gender equality – that females are perceived and raised to be inferior to men. The earliest evidence of these views (that we have uncovered) are written by Reverend James Fordyce in his *Sermons to Young Women* (1766) and *Addresses to Young Men* (1777). He instructs women to be submissive, meek, and sensitive, but men about the respect due to them, the reverence they owe themselves, and honor. Ahead of her time, Mary Wollstonecraft (1792) argues in *A Vindication of the Rights of Woman* that women’s rights should be equal to that of men. In doing so, she clearly highlights that at the end of the 18th century, men see women as “alluring mistresses [rather] than affectionate wives and rational mothers...”. She asks her fellow women to excuse her if she treats “them like rational creatures, instead of flattering their fascinating graces, and viewing them as if they were in a state of perpetual childhood, unable to stand alone.” An undated 19th century verse titled *Woman’s Rights* by M.C.M.R. makes clear that a woman’s place is in the home. The verse characterizes women as having the right to ‘train the infant mind’, ‘guide the tiny feet’, ‘solace the distressed’, ‘shelter the oppressed’, ‘be a bright sunbeam in high or lowly home’, ‘smile with loving gleam’, ‘fan the fevered brow’, and ‘to comfort man on earth and smooth his path to heaven.’

In the mid-19th century, the view of female inferiority persists, as Sara Stickney Ellis writes in (approximately) 1845 in *The Daughters of England* that “as women, then, the first thing of importance is to be content to be inferior to men”. John Ruskin’s 1865 lecture clearly delineates the separate characters of men and women, where man is “the doer, the creator, the discoverer, the defender” while “the woman’s power is for rule, not for battle, -- and her intellect is not for invention or creation, but for sweet ordering, arrangement, and decisions.” He further 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...”

In sharp contrast to that are mid-1800s calls for gender equality in Mill’s (1861) *The Subjection of Women* and Smith’s (1857) *Women and Work*, marking the beginnings of feminist activity. In arguing for equality, both works highlight the current inequality. Smith argues that it is degrading for a woman to be financially dependent on her husband. At this time (and until 1870), any property or money owned by a woman was controlled by her husband. Mill writes “That the principle which regulates the existing social relations between the two sexes – the

legal subordination of one sex to the other – is wrong in itself...”.

Though the view of women as the weaker, inferior sex is constant, some perceptions did change, most prominently the characterization of female sexuality. While the 18th century woman was thought to be ruled by her emotions and body, the 19th century woman was meant to be the mother and angel of the household, expected to stay chaste until marriage.¹⁷ With this shift came increased concern about prostitution, as evidenced by police records of the number of prostitutes and brothels. In 1857, the Met estimated there to be 8600 prostitutes and 2825 brothels (Acton, 1857). Such attitudes towards female sexuality are unlikely to largely influence our analysis since (i) prostitution is not trialed at the Old Bailey, (ii) we exclude gender specific offenses (including most sex offenses), and (iii) consider the within-offense gender gap.¹⁸

Women had no formal rights in politics in the 1700s and most of the 1800s. The feminist movements of the 1850s and 60s, including the Women’s Suffrage Movement starting in 1866, eventually resulted in an increasing number of rights: the right to vote¹⁹, unemployment benefits, rights with respect to marriage and divorce, and rights to be part of the legal profession and accountancy. Yet, it is not clear that the underlying gender attitudes changed dramatically. At the very least, some of English society still held these views in the early 1900s. A 1926 Daily Express newspaper article (five years after females began to sit on juries) says “Can woman, in short, suddenly divorce herself from temperamental inconsequence and from the compassionate and emotional instincts on which she often acts, and become in a moment a hard, matter-of-fact analytical administrator of the strict letter of the law? Many will be found to doubt it.”

3. Data: The Proceedings of the Old Bailey (1715-1913)

3.1. Data Description and Summary Statistics

Our data come from *The Proceedings of the Old Bailey*, which are a published account of all trials at each court session. The Proceedings were published from 1674 until 1913, but not consistently recorded until 1715. They initially provided entertainment for the population, with detailed transcripts of the most colorful cases. During the late 1700s, the Proceedings gained quasi-official status, being subsidized by the City of London government. At that point, all trials

¹⁷ See Old Bailey online discussion of gender roles and sexuality during this period, <https://www.oldbaileyonline.org/static/Gender.jsp#reading>, as well as a review by Kathryn Hughes in 2014, <https://www.bl.uk/romantics-and-victorians/articles/gender-roles-in-the-19th-century>.

¹⁸ In addition, such negative attitudes towards female prostitutes should lead to less lenient treatment of females by the courts – which works against our main findings.

¹⁹ Women over 30 were given the right to vote in 1918, but there were not equal voting rights between men and women until 1928.

were covered approximately equally, as the City demanded a ‘true, fair, and perfect narrative’.²⁰

We obtained xml files for about 2000 court sessions from 1715 to 1913 that were digitized by *The Old Bailey Proceedings Online*.²¹ From these files, we extracted ‘tagged’ information identifying the unique case, session date, defendant’s name, gender and age (generally only available for convicted defendants in the latter half of the sample), the offense as well as the verdict (plea, guilty of original or lesser charge, acquit) and sentencing outcomes (death, transportation, prison, corporal punishment, miscellaneous or no punishment). We can link all co-defendants to the same unique case, a data feature we take advantage of when testing the mechanisms. We manually coded judge, jury and juror names from 1750 to 1822. These are listed in the front matter to each Proceeding but not easily matched to the specific cases through the xml tags. We also manually coded defendant criminal history information that is available from the 1830s but not tagged. Criminal history indicates whether the defendant has been in custody once before (from 1832), more than once (from 1839) or is a known associate of bad character (from 1835). See Bindler and Hjalmarsson (2017) for additional details about the data.

Appendix Table 2 lists the 36 detailed offense categories in the initial data and the number of observations and share of female defendants in four periods spanning our sample: 1715-1750, 1750-1800, 1800-1850, and 1850-1913. We drop 13 offenses (23,735 out of 226,304 observations) for one of three reasons. First, some offenses had too few observations to be meaningfully analyzed. Second, to study the within offense gender gap, we drop offenses that are distinctly male (animal theft, embezzlement, mail theft, rape, sexual assault, and sodomy) or female (infanticide). For these offenses, the share of female defendants is generally less than 5% (or more than 95%) in almost all periods and the nature of the offense is often gender specific. Third, we drop observations with offense types ‘missing’ or ‘other’, which contain a wide range of not necessarily comparable offenses. Our final sample includes 23 detailed offense categories, which we categorize as: property, violent, fraud and other crime.²² As in Bindler and Hjalmarsson (2017), we also identified the capital eligibility of each offense. We further code whether the punishment received by a convicted defendant is the harshest punishment available at that time for a given offense.²³

²⁰ One exception is the period from 1790 to 1792 when only convictions were reported in the Proceedings. We exclude these years from our analysis sample.

²¹ <http://www.oldbaileyonline.org>

²² 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.

²³ For offenses that are capital eligible, the harshest punishment is the death penalty. For noncapital offenses, the harshest punishment is transportation when it exists and imprisonment otherwise. During two periods (the direct

The resulting data set consists of 202,569 trials over this approximately 200-year period. A better understanding of the data and criminal justice trends during this period can be obtained from Figure 1. Panel A presents the average annual number of cases by decade and broad offense category (property, violent, fraud, and other). It illustrates that property crime is the largest crime category throughout the time period. There is a large increase in the number of cases during the first half of the 1800s, which is due to reforms extending the catchment area of the Old Bailey (as mentioned earlier). The subsequent fall in the annual number of crimes can to some extent be attributed to jurisdictional changes shifting some of the less serious offenses to the lower courts (see King, 2006). Panel B of Figure 1 demonstrates that the share of female defendants decreased dramatically over this period, especially for property offenses: 38% of cases prior to 1750 had female defendants compared to 28% from 1750-1800, 22% from 1800-1850, and 13% from 1850-1913.²⁴ Panel C shows the share of guilty pleas and jury convictions by decade. Pleas were almost non-existent until the 1800s. But, by 1850, 29% of defendants plea; this increase primarily coincides with the shift in the burden of proof away from the defendant. Conviction rates were relatively stable at around 60% and increased during the 1800s, coinciding with the abolition of capital punishment. Finally, Panel D shows sentences (conditional on conviction) by decade and category. It clearly illustrates the shifts in sentencing regimes. Broadly speaking, capital punishment and transportation were common until the 1800s, when they were replaced by imprisonment.

Table 1 provides summary statistics for the whole period (1715-1913) and three sub-periods (1715-1800, 1801-1860, 1861-1913) by defendant gender. Note that these three periods can be characterized as before, during and after the 19th century justice reform period. During the entire period, approximately 20% of cases were capital eligible. However, prior to 1800 (pre death penalty abolition), 51% of male and 45% of female defendants were charged with capital offenses compared to just 1.2% and 2.7%, respectively, after 1860 (post death penalty abolition). Though 73% of all defendants are charged with property offenses, there again is a dramatic shift over time. In the first two periods, 71% and 83% of males and 81% and 86% of female are charged with a property crime. But, after 1860, these numbers drop to just 41% and 31%, respectively. This likely reflects the declassification of offenses as non-felonies (when they also become non-capital) and the shift of these now relatively more minor offenses to

aftermath of the American Revolution 1781-1786 and the stage-wise abolition of transportation 1853-1857) there is uncertainty about the harshest punishment available and we hence code it as missing.

²⁴ The share of female defendants in this last period is approaching that which is observed in criminal trials today. See Campaniello (2014) for a discussion of women in crime today.

lesser courts. There is thus a relative increase in the share of violent offenses observed in the data over time.

In terms of verdicts, Table 1 shows that 73% of defendants were found guilty by either pleading (14%) or a jury decision (58%). For males and females, the comparable statistics are 75% and 65%; that is, there is a 10-percentage point gender gap in the conviction rate. This raw gap is seen for each time period: 63% male vs. 55% female (pre-1800), 78% vs. 71% (1801-1860), and 78% vs. 67% (after 1860). Similar gender gaps are seen when considering just guilty jury verdicts for any offenses (70% males versus 62% females) or guilty jury verdicts of the original charge (62% males versus 50% females).

In terms of sentencing, over the entire period, 6.6% of guilty defendants were sentenced to death, 28.5% to transportation, 53% to prison and 4% to corporal punishment. In the pre-reform period, when capital punishment (the most severe punishment) was prevalent, 22% of male defendants versus 9% of females were sentenced to death. In contrast, females are more likely to be sentenced to transportation (56% vs 52%) or corporal punishment (15% vs 9%). Put differently, over the entire sample period, males are more likely to be sentenced to the harshest punishment available (57% vs 42%).

3.2. Raw Gender Gap in Jury Convictions and Plea Behavior

Before turning to a formal regression analysis, we graphically depict the raw (unconditional) gender gap over time. Figure 2 presents the share of guilty jury verdicts (of any offense) in Panel A and the share of pleas (Panel B) for male and female defendants by decade. Throughout these 200 years, females are less likely to be convicted by the jury, with the gap ranging from 5-15 percentage points. Once pleading was integrated into the legal system in the 1820s, males are persistently more likely to plead guilty – the jury (Panel A) and the defendant (Panel B). However, a defendant may plead because of an expectation of how likely the jury is to convict in his/her case. This expectation could differ by defendant gender, especially if defense attorneys assist in this decision. Given the different agents, we conduct most of the analysis separately by decision/agent. However, we also present robustness specifications that treat guilty pleas as if they are guilty jury verdicts. In that spirit, Panel C of Figure 2 shows the overall gender gap in convictions by any means (jury or guilty plea). Not surprisingly, adding the gap in plea behavior amplifies the overall gap in the latter half of the sample (except for the last decade, which is based on just three years and excluded from the regression analyses).

Of course, these raw (unconditional) gender gaps can result from a wide range of

explanations, including, most simply, a differing distribution of offenses (with different baseline conviction rates) by gender. Though our regression analysis considers this more formally, we provide suggestive evidence here that this simple explanation is not underlying the conviction gender gap. Figure 3 presents the raw gender gap separately for each of the largest offense categories (larceny, theft from place, pick-pocketing, burglary, coining, robbery, receiving, and fraud). Within each offense category and persistently throughout the period, females are less likely to be convicted than males.

3.3. Raw Gender Gap in Judge Sentencing

Figure 4 presents the share of convicts sentenced by the judge to capital punishment, transportation, prison, corporal punishment, miscellaneous or no punishment (panels A to F, respectively) by gender. A few patterns stand out. Throughout the pre-abolition of capital punishment period, females are less likely to be sentenced to death; the gap gets smaller as the reform period (primarily 1810-1850) progresses. The share sentenced to transportation is higher for females before the Revolution, i.e. when transportation was the main alternative to capital punishment. After the halt of transportation and the emergence of prison as a more lenient sentencing alternative, males are sentenced to transportation more than females. This is also true when capital punishment is abolished and transportation is the harshest penalty available. The patterns for imprisonment are in line with this: females are more likely to be sentenced to prison than males when harsher penalties are at stake, but less likely from the mid-1800s when both capital punishment and transportation were largely abolished. For corporal punishment, similar patterns can be seen although at a much lower scale. Miscellaneous and no punishment shares are generally higher for females than for males. These figures highlight the importance of considering which punishments were available at a particular time: the direction of the raw gender gap for a sanction depends on the sanction's relative severity, which changes over time.

As is the case for conviction, gender differences in sentencing can simply reflect different distributions of offenses, including whether the offense is capital or not. To explicitly illustrate that (i) gender gaps in sentences exist within offense category and (ii) how sentencing gender gaps change as expected punishment (i.e. the harshest sanction) changes, Figure 5 presents shares of convicted males and females sentenced to each sanction for burglary.²⁵ Burglary was capital until 1837; transportation then became the most severe potential punishment and upon its abolition in the 1850s, prison became the harshest punishment. A clear pattern emerges:

²⁵ Comparable figures for any offense category or sentence are available from the authors upon request.

females are generally less likely to receive the most severe punishment available than males. Before 1837, there is a large visible gap (favoring females) in the share of burglary offenders receiving death sentences. Rather, females were more likely to be sentenced to transportation or prison (once an option). When transportation is abolished, and prison is the harshest sanction, the gender gap in prison reverses; females become less likely to receive a prison sentence and more likely to receive a miscellaneous or no punishment.

As a way of summarizing the gender gaps for the various sanctions, and to ease the presentation of the following regression analyses, we present the raw gap in the likelihood of being sentenced to the harshest punishment available at the time for a given offense, i.e. taking into account the capital status of the offense. Panel A of Figure 6 demonstrates in the raw data that females are persistently less likely than males to be sentenced to the harshest punishment available. One can also see in this figure that judges were much more likely to hand out the harshest punishment (to both males and females) after the abolition of transportation; that probability jumps from levels of around 40% in 1859 to more than 90% in 1860. Despite this dramatic change in judge behavior, the gender gap persists.

4. Regression Analysis

4.1. The Gender Gap and Observable Case Characteristics

The simplest explanation for the raw gender gaps is that they are an artefact of a differential distribution of offenses and case characteristics (with different conviction rates or potential punishments) by defendant gender. This section formally tests this explanation: We estimate the gender gap in the verdict or sentencing outcome net of observable case characteristics.

$$Outcome_{it} = \alpha + \beta Female Defendant_i + X_i + \gamma_t + \varepsilon_{it}$$

The baseline set of observables (X) includes the number of defendants, the 23 detailed offense type dummies, and whether the offense is capital. As criminal history is only available after 1830, we do not include it in the baseline, but present robustness checks for the periods during which it is available. Note that even in this historical period, our controls are similar to that of many modern empirical papers. The baseline specification includes year fixed effects, which capture unobservable characteristics of, for instance, the justice system common to male and female defendants; our results are, however, robust to the exclusion of the year dummies.

We begin by estimating the above specification by decade and graphically depicting the

regression adjusted gender gap and associated confidence interval in Figure 7 for conviction and pleas. Panel A of Figure 7 suggests that little to none of the raw gender gap in convictions is explained by differential case characteristics. In fact, the estimated gap fluctuates around 10 percentage points for the entire 200-year period and remains significant throughout. In other words, the within offense category gender gap is as large as that across offenses. Panel B shows that females are 5-10 percentage points significantly less likely to plead guilty in the latter 19th century (once pleading occurs regularly), even when controlling for observables. The findings from Panel A are also robust to pooling jury convictions and guilty pleas, as shown in Panel C.

With respect to sentencing, Figure 8 tells the same basic story as the raw gender gaps: observable differences in case characteristics do not explain the gender gaps. Thus, even in the conditional estimations, females are significantly less likely to receive a death sentence (when capital punishment existed) and more likely to be sentenced to corporal or miscellaneous punishment during this period. When prison was possible (after the first prison ships of the Revolution), females were less likely to be transported and more likely to be sentenced to prison than males. But, once transportation was abolished and prison became the harshest sanction, the gap switches in sign and females are less likely to be sentenced to prison. Similarly, one can see this when looking at being sentenced to the harshest available sentence in Panel B of Figure 6: Females are persistently less likely to be sentenced to the harshest punishment available with the largest gaps during the periods that did not see major reforms in the criminal justice system.

Table 2 presents the regression results for conviction and pleas in tabular format, pooling the data into approximately 50-year periods. Column (1) presents the raw gap with no controls while column (2) presents the gender gap adjusted for observable controls. Panel A looks at whether the jury convicts the defendant of any charge: adding the observable controls actually has no effect or generally slightly increases the magnitude of the gender gap. Compared to the pre-1751 period when females were 12 percentage points (or 20.3%) less likely to be convicted, the gender gap in later periods is smaller though still large and significant (0.08 or 13.5% in 1751-1800, 0.07 or 9.4% in 1801-1850, and 0.09 or 14.1% after 1850). Panel B considers an alternative dependent variable – whether the jury convicted the defendant of the initial (most serious) charge. The raw gender gap here is generally larger than that for any conviction and again robust to observable characteristics. This suggests that juries treat female defendants favorably by convicting them of a lesser charge. Panel C considers pleas: with the full set of observable controls (column (2)), females are about four percentage points less likely to plead

guilty from 1801-1850 and six percentage points after 1850.²⁶

Table 3 presents comparable estimates for whether the defendant was sentenced to the harshest punishment available at the time. Columns (1) and (2) consider all offenses together, excluding and including observable controls, respectively. Given the importance of the punishment regime, Columns (4) and (6) present the results for capital and non-capital offenses, respectively, including the full set of controls. The bottom line is that large and significant gender gaps in sentencing are seen, even after controlling for observables: Females are less likely to be sentenced to the harshest punishment available at the time. For all offenses, the gender gap ranges from 4.2 percentage points (after 1850) to 10.6 percentage points (1751-1800). The gaps in the likelihood of receiving the harshest punishment in each period are reasonably comparable for capital and non-capital offenses. It is worth highlighting that the gender gap in sentencing is much smaller after the abolition of the death penalty and transportation (i.e. after the 1850s), when prison became the harshest punishment. While this could be due to the relative severity of punishments, it could also be that the gender gap in prison is amplified at the intensive margin – sentence length – which we do not observe.²⁷

As a way of summarizing the importance (or lack thereof) of differences in observable characteristics in explaining the gender gap in convictions and sentencing, Table 4 presents a Blinder-Oaxaca decomposition for (any) conviction and being sentenced to the harshest punishment available. That is, we decompose the gender gaps into the share that can be explained by differences in observable characteristics as opposed to the share due to differences in the coefficients on those characteristics. Columns (1) to (3) show the decomposition for the whole sample where we gradually include more control variables, starting with year fixed effects only, adding detailed offense type dummies and then adding controls for whether the offense is capital and the number of defendants. Using the specification with the full set of baseline controls, columns (4) – (7) look at different time periods (before 1751, 1751-1800, 1801-1850, after 1850). For the last subsample (after 1850), we are able to control for criminal history (column (8)) over and above the baseline controls and in the sentencing sample also for the age of convicted defendants (column (9)).²⁸

²⁶ These results are completely robust to treating pleas as guilty jury verdicts. Results are available upon request.

²⁷ Separate regression results are available upon request for each punishment (death penalty, transportation, prison and corporal punishment) and are consistent with the results shown here for our summary measure of the harshest punishment available: Females are 5-10 percentage points less likely to be sentenced to death for capital cases, 6-12 percentage points less likely to be sentenced to transportation for non-capital cases when transportation was on the table, and two percentage points less likely to be sentenced to prison for non-capital offenses once transportation is off the table.

²⁸ From around 1830 onwards, criminal history is available as an additional observable control. Though not available for earlier periods, one should not consider it ‘unobserved’ in these periods: since most convicts were

For convictions, the gap widens once we control for the offense type; adding additional control variable does not explain that gap. For the sentencing outcome, part of the gender gap (34%) is accounted for by variation over the years. That share increases to 47% once we control for offense types. This is unsurprising given changing sentencing regimes over that 200-year period. Adding additional controls does not explain the gender gap; this is even true when we control for criminal history and defendant age. The bottom line is that these observable differences explain little of the raw gap overall and in any of the sub-periods: Pretty much the entire difference in the gender gap in convictions and a large part of the gender gap in sentencing is driven by differences in the coefficients rather than by differences in the characteristics.

How do these findings compare to analyses of the gender gap in the labor market? If one draws a parallel between occupation/industry and crime type/category, then our findings contrast the conclusions of Blau and Kahn (forthcoming) concerning the wage gender gap today. Whereas industry/occupation explain a significant portion of the wage gender gap, crime type explains little to no of the criminal justice gender gap. On the other hand, the majority of the gender wage gap is left unexplained by differences in observable characteristics including industry and occupation (Altonji and Blank, 1999).

5. Unobservables as a Potential Explanation for the Unexplained Gender Gap

This section focuses on discussing and – where possible - testing explanations for the gender gap that can be linked to unobservable case characteristics, including the female as better defendant and the assignment to more lenient judges and juries.

5.1. Unobserved Differences in Case Characteristics

Though conditioning on observables had little impact on the estimated gender gaps, this does not rule out the possibility that there are unobservable differences in case characteristics by gender. For instance, even if charged with the same offense category, the offenses for female defendants may be systematically less severe; for property crimes, this may involve goods of lesser values while, for violent crimes, this may involve fewer victims. This data is not systematically tagged in the proceedings, but would be available to the jury and judge. That

killed or transported, recidivism did not really exist as a phenomenon. In contrast to today, during this historical period, criminal history is information that is presented to the jury; moreover, with increased tracking of criminals and concerns about recidivism, being a ‘habitual criminal’ (someone who has reoffended three times) was classified as a criminal offense in and of itself in 1908. An additional observable control available for the sentencing outcomes after 1800 is age; it cannot be used for conviction regressions since it is only consistently reported in the proceedings for defendants who are found guilty.

said, we believe the extent to which unobservable case characteristics, including evidence quality, is a concern is limited, given that minimal evidence was actually presented at trial. Feeley (1997) describes shockingly short trials (on average, eight minutes in the early 1800s) that are more consistent with the modern day sentencing phase of a trial.

Yet, to empirically test whether unobserved differences explain a part of the gender gap, we use the number of words per trial printed in the Proceedings as a proxy for unobserved characteristics. We recorded the number of words for every reported trial in the February, May and September sessions from 1751 to 1810.²⁹ Regressions (available upon request) 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. We argue that juries may have additional information (e.g. number of victims, testimony length, witness credibility or sensationalism) that is not observed by us but, given that testimony is often published verbatim, likely to be captured by word count.

For the word count sample, Table 5 presents the raw gender gap in column (1), the observables adjusted gap in column (2), and the observable and unobservable (word count) adjusted gaps in columns (3) and (4). Panels A and B present conviction of any charge and initial charge, respectively, while Panel C shows the harshest punishment outcome.³⁰ The same pattern of results is seen as in the full 1751-1810 sample. Controlling for word count – either linearly (column (3)) or more flexibly with word count quartile dummies (column (4)) – over and above the other observables has little to no effect on the gender gap. Thus, it is highly unlikely that the gender gaps can be explained by unobservable case differences.

5.2. Females as More Successful Defendants

Another potential explanation for the pro-female gender gap is that females were more successful defendants. To some extent, this can be thought of as another variable that is unobserved to us, but observed by the jury. That is, females could have been better at defending themselves, perhaps by providing more convincing testimony or better witnesses. Though this channel is hard to test empirically, we argue that it is unlikely to be the driving mechanism for two reasons. First, the quality of the defense, potentially measured by the number of witnesses and length of testimony, would be at least partially captured by the number of words printed in the Proceedings. But, we have already seen that the gender gap is completely robust to word

²⁹ Word count is collected for these years because these years have jury identifiers. Bindler and Hjalmarsson (2017b) use the word count as a measure of unobservables to study path dependency in jury decision making.

³⁰ Results for the four separate punishment outcomes are available upon request.

count. Second, given that trials during this period were quite short and that it was difficult to enforce the attendance of witnesses due to the irregular and unknown scheduling of the session, it is hard to imagine how one would distinguish themselves as a ‘good’ defendant.

5.3. Within Judge and Jury Gender Gaps

Another potential unobservable is that females are systematically assigned to judges and juries who are more ‘lenient’ in their treatment of all defendants. As is the case today, there is substantial variation across juries in the share of defendants convicted, and across judges in the share of defendants sentenced to maximum punishment, as illustrated in Appendix Figure 2. While we do not have any anecdotal evidence of such non-random assignment, this section empirically rules out this channel by estimating within judge and jury gender gaps. While we cannot observe judge and jury identifiers for the entire 200-year period, we can include all judge and jury fixed effects for the subsample 1750-1822, when such identifiers are available. This precludes the inclusion of year fixed effects, which are perfectly collinear with the jury dummies. The estimated gender gaps are quite robust to the inclusion of jury and judge fixed effects for jury verdict and plea behavior in columns (3) and (4), respectively, of Table 2. This is especially true for the 1750-1800 period, for which we know the judge and jury for almost all observations. The same holds when adding judge and jury fixed effects to the sentencing regressions for all cases (column (3)), capital cases (column (5)) and non-capital cases (column (7)) in Table 3.³¹

6. Differential Attitudes and Discrimination as an Explanation of the Gender Gap

Our conclusions from the above analyses suggest that unobservable differences in case characteristics (including defendant ability and judge and jury assignment) are unlikely to explain much of the gap. This leaves differences in attitudes towards female and male defendants as the remaining channel. Such differential attitudes can arise for multiple reasons. One possibility is the idea that females are simply not capable of actually being ‘responsible’ for the crime, which is formalized in the girlfriend theory. Another possibility is that male judges and juries treat females more leniently out of a desire to protect children by leaving a caretaker in the home. Alternatively, these differences in attitudes and perceptions of females

³¹ A large number of observations is lost when adding judge and juries from 1800-1850 since they are only available for the first half of the period. Any differences in estimates (e.g., the sentencing gender gaps appear somewhat larger with judge and jury) could result from different samples; this is especially likely given that the reforms of capital punishment primarily occur just after the end of the judge and jury data.

may result in lenient treatment of females (or harsher treatment of males) because of discrimination. We will discuss the possibility of both statistical and taste-based discrimination.

6.1. Females Not Capable of Crime - The Girlfriend Theory

Starr (2015) puts forward the girlfriend theory, where decision makers deem it not possible for a female to be responsible of a crime; someone else must be responsible. This channel is a formalized part of the legal system at the time for married couples with *feme covert*. According to the Old Bailey Online, this principle implied that “women could not be held responsible for crimes committed in the presence of their husbands (since they were presumed to be following their husbands’ commands)”.³² Though it was not often officially applied by juries (and is not a verdict subcategory in the data), it could be that the gap arises from an informal application of this principle by the jury in mixed gender pairs.

We examine the extent to which this mechanism can explain the historical gender gap in jury verdicts in Table 6. Specifically, we re-estimate the adjusted gender gap in the chance of conviction for single defendant cases (column (1)) and multiple defendant cases (column (2)). We then divide the multiple defendant cases into subsamples of same gender (column (3)) and mixed gender (column (4)) cases. Column (5) adds case fixed effects to the mixed gender sample, and explicitly tests whether female defendants are treated more leniently than males in the same case. Columns (6) and (7) show mixed gender pairs of defendants who do and do not have the same name respectively; this is our best indication of marriage, though we may be identifying other relationships (e.g. siblings). The first row presents the results for all offenders (with controls) while the remaining rows present the results by offense category. We show property, violent, and fraud offenses with more than 250 observations of mixed gender multiple defendant cases (to allow for a test of the girlfriend theory).

What can this analysis tell us about the mechanisms? First, examining the single defendant gender gap is useful in itself. If there still is a gender gap in convictions, then ‘the girlfriend theory’ cannot be its sole explanation. Second, a comparison of the mixed and single gender multiple defendant cases indicates whether the girlfriend theory plays any role: if the mixed gender estimates are greater than the single gender estimates, then this supports the girlfriend theory. Finally, the ‘girlfriend theory’ is most likely applicable to married co-defendants; one would expect the largest effects for same name mixed gender co-defendants.

The first important finding in Table 6 is that females are six percentage points (8.5% at the

³² See <https://www.oldbaileyonline.org/static/Gender.jsp>, sourced on August 17, 2017.

mean) less likely to be convicted than males when looking at all single defendant cases. In addition, a significant negative and substantive gender gap (ranging from -0.05 for larceny to -0.20 for pickpocketing) is observed for the single defendant sample in almost every offense category. This suggests that the girlfriend theory is far from being the sole explanation for the conviction gender gap. In fact, for an offense such as robbery, it likely plays little to no role since all of the gender gap is driven by the single defendant cases. The second finding, however, is that there is some support of the girlfriend theory as a mechanism for a subset of crime categories. Specifically, the mixed gender multiple defendant estimates are larger than the single gender estimates for burglary, housebreaking, receiving, stealing from master, coining, and fraud. In addition, for every offense category, the gap is largest for the same name pairs compared to different name pairs. This is even true for offenses for which the mixed gender gap (overall) was not larger than the single gender gap. Though there is some support of the ‘girlfriend theory’, this can only explain some of the gap in a relatively small number of cases.

6.2. Female as Mother – A desire to protect the children

Another possibility is that females are treated more leniently by the courts because of their role in the home. That is, judges and juries may be trying to protect the children of these defendants by not leaving them ‘motherless’. While we cannot systematically observe the number of children of each defendant, we provide some suggestive evidence regarding this channel.

Specifically, we take advantage of the Old Bailey Corpus Online, which provides a search engine of all spoken text in the Proceedings. When possible, it also identifies the type of speaker and their gender. As there is very little spoken text of the judge and jury in the Proceedings, we cannot explicitly observe whether children factored into their decisions. However, we can at least look at the statements made by the other parties in the courtroom – especially defendants, victims, and witnesses. Specifically, we find 2,154 utterances of the word ‘children’ (in just over 1000 trials); 57% were by witnesses, 10% by defendants, 14% by victims, and 13% with an unknown speaker role.³³ Given that there are more 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. With regards to whether children differentially affected outcomes for male and female defendants, we look directly at the number of utterances of ‘children’ by defendant gender.

³³The 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 the speaker role or year. Note that we did not search on alternative words like ‘child’ because this is often used at the time in an alternative context – to be as innocent as the child unborn. Searches on other words for children were considered but found not useful – kid (handful of hits related to leather) and kids (no hits).

39% of the utterances are by female defendants. While this is greater than the rate at which female defendants are observed in the data, it certainly does not appear to be a channel that is gender specific. Many of the utterances of children for male defendants were of the form: ‘I have a wife and N small children.’ Moreover, the overwhelming majority of defendants do not mention ‘children’. Another argument against this channel is that the male is often the ‘breadwinner’ of the household. Thus, protecting the child would be just as important when considering the possibility of removing the father from the household. This idea is consistent with tabulations by defendant gender above.

A potential alternative test for the importance of this channel is to look at subsamples by defendant age, to the extent that an individual’s age is correlated with having young children at home. Since this is only systematically reported for convicted defendants after 1800, we can only look at the sentencing stage for the latter part of our sample (1801-1850 and 1850 to 1900). Controlling for defendant age in ten-year intervals has no effect on the gender gap. Further, results obtained from specifications that allow for a differential gender gap in each age bracket are not consistent with a ‘female as mother’ channel driving the gender gap.³⁴

6.3. Statistical Discrimination

This section considers whether statistical discrimination plays a role: are juries less likely to find females guilty because of a prior belief that few(er) criminals are female? The Proceedings were published for public consumption, and almost all juries contained jurors who had previous jury experience. Thus, it seems plausible that the juries would have a sense of the share of female defendants (a proxy for criminals). To empirically test this channel, we consider two features of the data. First, there was a substantial change in the share of female defendants at the Old Bailey over time: it decreased from almost 40% to almost 10% (close to modern-day levels of criminal participation) over the time period. Second, there is substantial variation across offenses in the share of female defendants.

We begin with the observed decline in female criminals at the Old Bailey. There is, in fact, some debate about why such a decline occurs. Some historians argue that it reflects a true decrease in female criminality, as the female has become the angel of the home with less opportunity to commit crime (Feeley and Little, 1991). If this is the case, then one should not

³⁴ Specifically, for 1801-1850, the gender gap is increasing with defendant age, whereas preferential treatment towards mothers would suggest the largest gaps for the 20-30 and potentially 30-40 age brackets. After 1850, the gender gap is largest for the juvenile defendants (under age 20), which could be driven by juvenile justice reforms at the time. Results available upon request.

expect there to be substantial change in the ‘type’ of female criminal observed at the Old Bailey. That is, if the share of female defendants can be simply taken as a proxy for the share of female criminals, then statistical discrimination would predict a widening gender gap over time. In other words, if juries use a defendant’s gender to infer about the likelihood of guilt, they should be even less likely to believe a female is guilty at the end of the period, when there are so few female ‘criminals’. On the other hand, if the explanation for the decreasing share of female defendants at the Old Bailey is because they are disproportionately shifted to lesser courts (King, 2006), then one might think that the quality of female criminal at the Old Bailey is worse; that is, though there are fewer females observed at the Old Bailey, these individuals represent the most criminal females. In that case, statistical discrimination would predict a narrowing of the gender gap over time.

The first suggestion that statistical discrimination did not play a prominent role is seen in Table 2 and Table 3 looking at gender gaps in convictions and sentencing in 50-year brackets, and Figure 6 to Figure 8 showing the regression adjusted gender gaps by decade. The gender gap in conviction, for instance, remains relatively constant over time: none of these figures demonstrate a gap that is continuously widening or narrowing over time.

Table 7 tests this more formally by estimating the conviction and sentencing gender gap (with controls) for subsamples differing in their share of females. The two panels correspond to the main conviction and sentencing outcomes, respectively. Column (1) looks at the subsample of offenses with a decreasing share of female defendants over time, while column (2) looks at those (few) offenses with an increasing share.³⁵ Instead of focusing on the variation in the share of female defendants over time, columns (3) and (4) zoom in on the (relatively more) cross-sectional variation by splitting the sample into those sessions in which the share of female defendants is above and below the median share in that 50-year bracket, respectively. That is, we consider the possibility that jury and judge perceptions on the criminality of females are determined by the share of female defendants that they directly observe in their session. Regardless of how we split the sample, we do not see a pattern in the estimated coefficients that would suggest a systematic widening or narrowing of the estimated gender gap, neither for convictions nor for sentencing. Generally speaking, the gender gaps are quite comparable for sessions with a relatively high or low share of female defendants.

³⁵ Offenses with an (unambiguously) decreasing share of female defendants include: burglary, coining, fraud, housebreaking, larceny, pickpocketing, receiving, robbery, shoplifting, stealing from master, theft from place. Offenses with an (unambiguously) increasing share of female defendants include: manslaughter, murder, wounding. For more details on the share of females by offense and over time, see Appendix Table 2.

Table 8 estimates the (adjusted) gender gap separately by offense, and ranks the estimates according to the share of female defendants in that category. The highest share of female defendants (46%) is seen for shoplifting. 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%). A statistical discrimination story suggests that the gender gap is larger for the least female offenses. Yet, we do not see such a systematic pattern. For the offenses listed here, the corresponding gender gaps are -0.04, -0.10, -0.19, -0.07, +0.03, -0.15, -0.07, and -0.21.

A final argument against statistical discrimination as a mechanism is that the gap is seen for both verdicts and sentencing. While one can imagine a statistical discrimination scenario for verdicts, it is harder to imagine for sentencing. The defendant has already been found guilty: there is no (at least, less) uncertainty left for which a signal given by gender may be useful. The conclusion of the above observations is that statistical discrimination seems an unlikely candidate as an explanation of the gender gaps seen in the data.

6.4. Preference Based Discrimination - Males as ‘Protectors of the Weaker Sex’

The remaining channel to consider is preference based discrimination, which one can think of as paternalism of the all-male judiciary taking on the role of protecting the so-called ‘weaker’ sex. Given that males were deemed responsible for the welfare of females (their wives) in the home, it certainly seems feasible that they carried this duty over to the courtroom. In the context of the traditional labor market taste-based discrimination model, one can think of judges and jurors as being less likely to convict females because of their positive taste/preference for protecting them. Does this have any testable implications? One is that juries are willing to acquit female defendants 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 females relative to men. Empirically, thus, the quality of evidence for convicted females should be greater than that for males. Unfortunately, we do not currently have any suitable measures of the quality of evidence at the case level, but are considering the possibilities of additional data collection.

Although we cannot yet provide an explicit empirical test of preference-based discrimination (as in Starr, 2015), knowledge of societal gender roles during this period makes it a highly plausible explanation for the gender gap. In particular, the persistence of the gap

coincides with the persistent attitudes towards females characterized earlier in the paper.

7. Conclusion

This paper finds a criminal justice gender gap favoring females that (i) is seen at multiple stages of the justice system, ranging from pleas to conviction and sentencing, and (ii) persists throughout two centuries of trials: Females were less likely to be found guilty, and conditional on guilt, less likely to receive the harshest punishment available. Why?

Oaxaca decompositions suggest that almost all of the raw gender gaps in conviction and sentencing are not explained by gender differences in observable case characteristics. We also demonstrate that unexplained gender gap is not driven by either differential assignment of females to more lenient judges and juries or unobservable case differences (including the female as better defendant), using the number of words published per trial as a measure of unobservables. We also rule out preferential treatment to female defendants because they are mothers with childcare responsibilities as a likely channel, as defendant testimony rarely states the number of children and does not do so at a dramatically different rate for male and female defendants. We further argue that the persistence of the gap rules out statistical discrimination; the gap does not systematically widen (or decrease) over time as the share of female defendants decreases. Finally, while the ‘girlfriend theory’ may explain some of the gender gap in multiple defendant cases, especially for spouses, this is not a possible explanation for the still existent gender gap in single defendant cases.

This leaves us with taste-based discrimination, which can be characterized as paternalism by the all-male judiciary, as a potential explanation. While we cannot (yet) empirically test this channel, we describe that the persistence of the gender gap over this time period is consistent with the persistence of societal attitudes towards females as the ‘weaker sex’ and in need of protection. This conclusion is in line with the modern-day findings of Schanzenbach (2005) and Philippe (2017) 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. (forthcoming), based on a rank order test, that tastes for gender discrimination are unlikely to be driving the gender disparity in sentencing.

How does our paper contribute to the understanding of today’s gender gap? 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? One common factor is that judges both today and then come from a different socioeconomic class and

background than most defendants. Even in the 18th and 19th centuries, many judges were still university educated. Perhaps more of a concern is the fact that the modern-day English (and American) jury pool is drawn from the general population, whereas the main eligibility criteria for jury service in our period is wealth-based. Anecdotal evidence, however, supports the notion that the educational standards and social class of jurors declined in the 18th and 19th centuries.³⁶ This is in part driven by a lowering of the wealth threshold for service in the Juries Act of 1825. Yet, despite the changing composition of the jury pool, the gender gap persisted throughout our sample, suggesting that attitudes towards women were not specific to a small class of society, and may therefore be relevant to the broader pool today.

During our 200-year sample, there is a prevalent and explicit bias towards women as the weaker sex. We argue that this bias is seen in the courtroom (to the benefit of women). While the writings of today do not exhibit the same explicit bias as those of the 18th and 19th centuries, these beliefs may still be prevalent today and carry over to the courtroom as an implicit bias. That is, (male and female) agents of the justice system today may be unconsciously and unintentionally discriminating (Bertrand et al, 2005) on the basis of gender. Thus, it may be that despite all of the ground gained by women, this perception has not 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 (2016) “A Jury of Her Peers: The Impact of the First Female Jurors on Criminal Verdicts” NBER Working Paper No. 21960.
- Beattie, J.M. (1986) *Crime and the Courts in England 1660-1800*, Princeton University Press, pp 663ff.
- Bentley, D. (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, A. and R. Hjalmarsson (2017a) “The Fall of Capital Punishment and the Rise of Prisons: How Expected Sentences Affect Jury Verdicts,” CEPR Discussion Paper No. 11888.
- Bindler, A. and R. Hjalmarsson (2017b) “Path-Dependency in Jury Decision Making,”

³⁶ See <https://www.oldbaileyonline.org/static/Judges-and-juries.jsp#reading>, last accessed September 4, 2017.

Working paper.

Blau, Francine and Lawrence Kahn (forthcoming) "The Gender Wage Gap: Extent, Trends, and Explanations," *Journal of Economic Literature*.

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 (forthcoming) "Comparing Apples to Oranges: Differences in Women's and Men's Incarceration and Sentencing Outcomes" *Journal of Labor Economics*, 35, S1, S201-S234.

Campaniello, Nadia (2014) "Women in Crime" *IZA World of Labor*: 105.

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.

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

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

Fordyce, James (1777) *Addresses to Young Men*. Volume I, The Second Edition, T. Cadell in the Strand, London.

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, P. (2006), *Crime and Law in England, 1750-1840*, Cambridge

M.C.M.R (undated 19th century) *Woman's Rights*.

Mill, John Stuart (1861) *The Subjection of Women*. Second Edition. Longmans, Green, Reader, and Dyer. London.

Mustard, D.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. (2017) "Gender Disparities in Criminal Justice" Toulouse School of Economics Working Paper 17-762.

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

Schanzenbach, M.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.

Smith, Barbara Leigh (1857) *Women and Work*. Bosworth and Harrison, London.

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, S. (2015) "Estimating Gender Disparities in Federal Criminal Cases." *Am. L. & Econ. Rev.* 17, no. 1: 127-59.

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

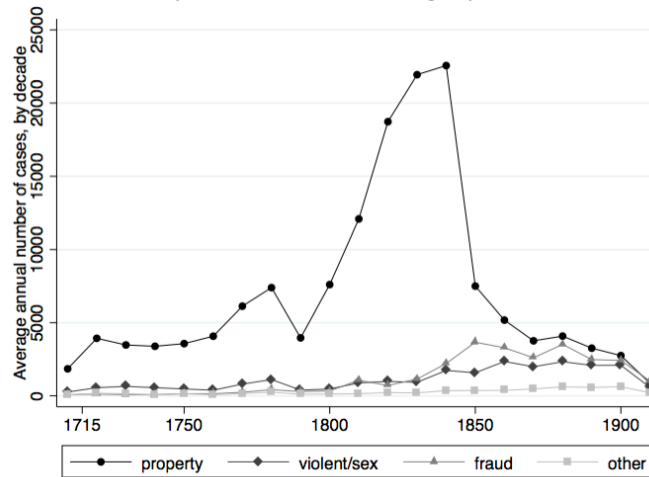
Stolzenberg, Lisa and Stewart D'Alessio (2004) "Sex Differences in the Likelihood of Arrest." *Journal of Criminal Justice*. 32: 443-54.

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

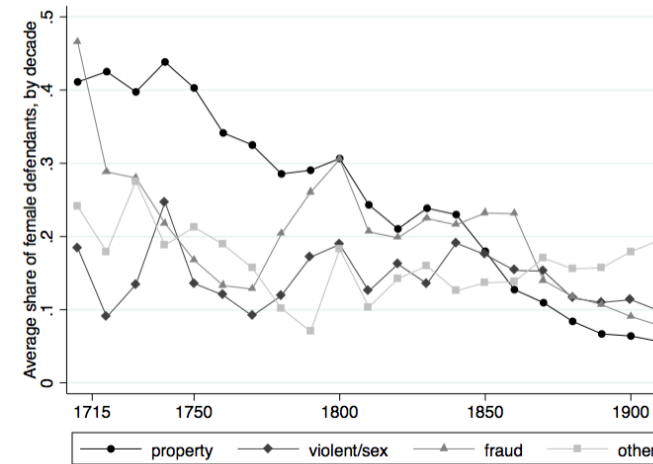
Wollstonecraft, Mary (1792) *A Vindication of the Rights of Woman: with Strictures on political and Moral Subjects*. Printed at Boston, by Peter Edes for Thomas and Andrews, Faust's statue, no. 45, Newbury-street, MDCCXCII. Pp. 452.

Figure 1. Number and Shares of Cases by Decade

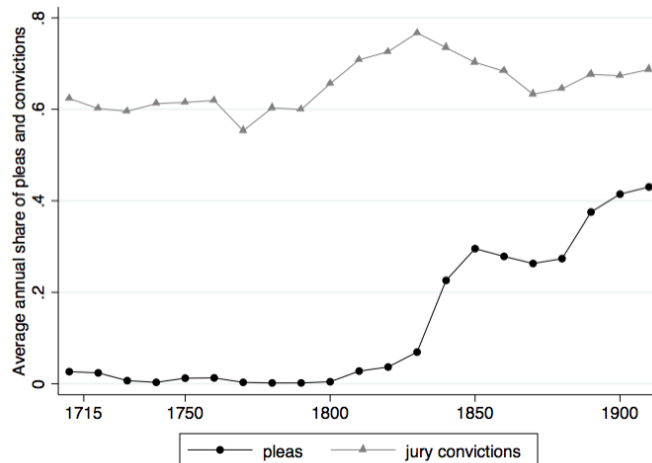
A. Number of cases by broad offense category



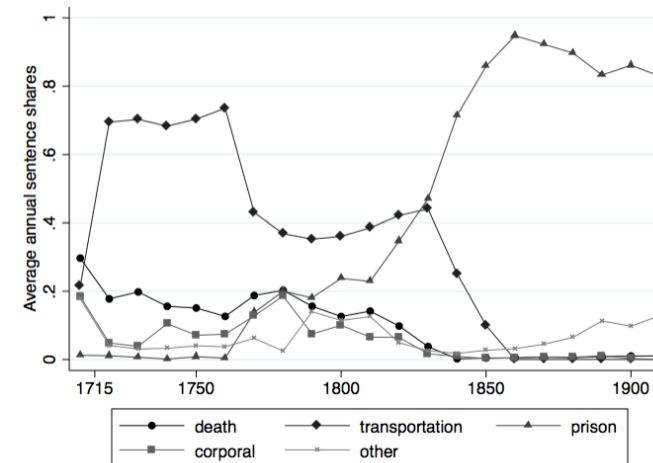
B. Share of female defendants by broad offense category



C. Share of pleas and convictions



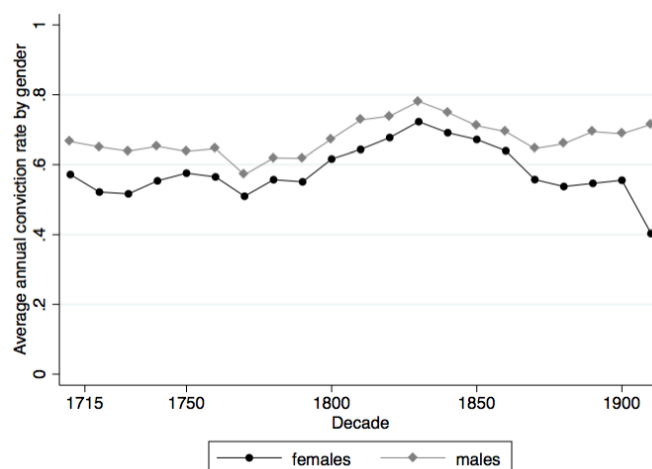
D. Share of sentence types



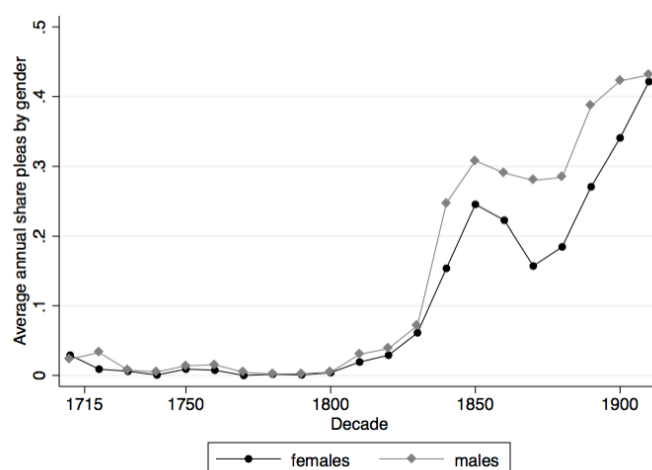
NOTE – Figure A shows the average annual number of cases (averaged by decade), Figure B the average share of female defendant cases, Figure C the share of cases with pleas or convictions and Figure D the share of cases by different sentence types (conditional on conviction). SOURCE - *The Old Bailey Proceedings Online* and own calculations.

Figure 2. Raw Gender Gap in Jury Convictions and Guilty Pleas

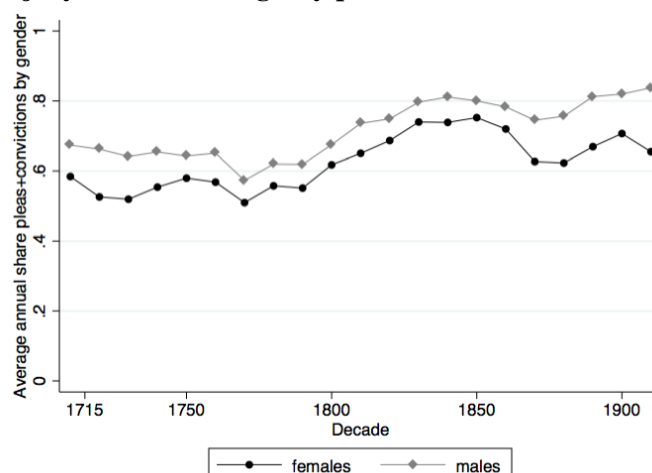
A. Raw gender gap in jury convictions of any offense



B. Raw gender gap in guilty pleas

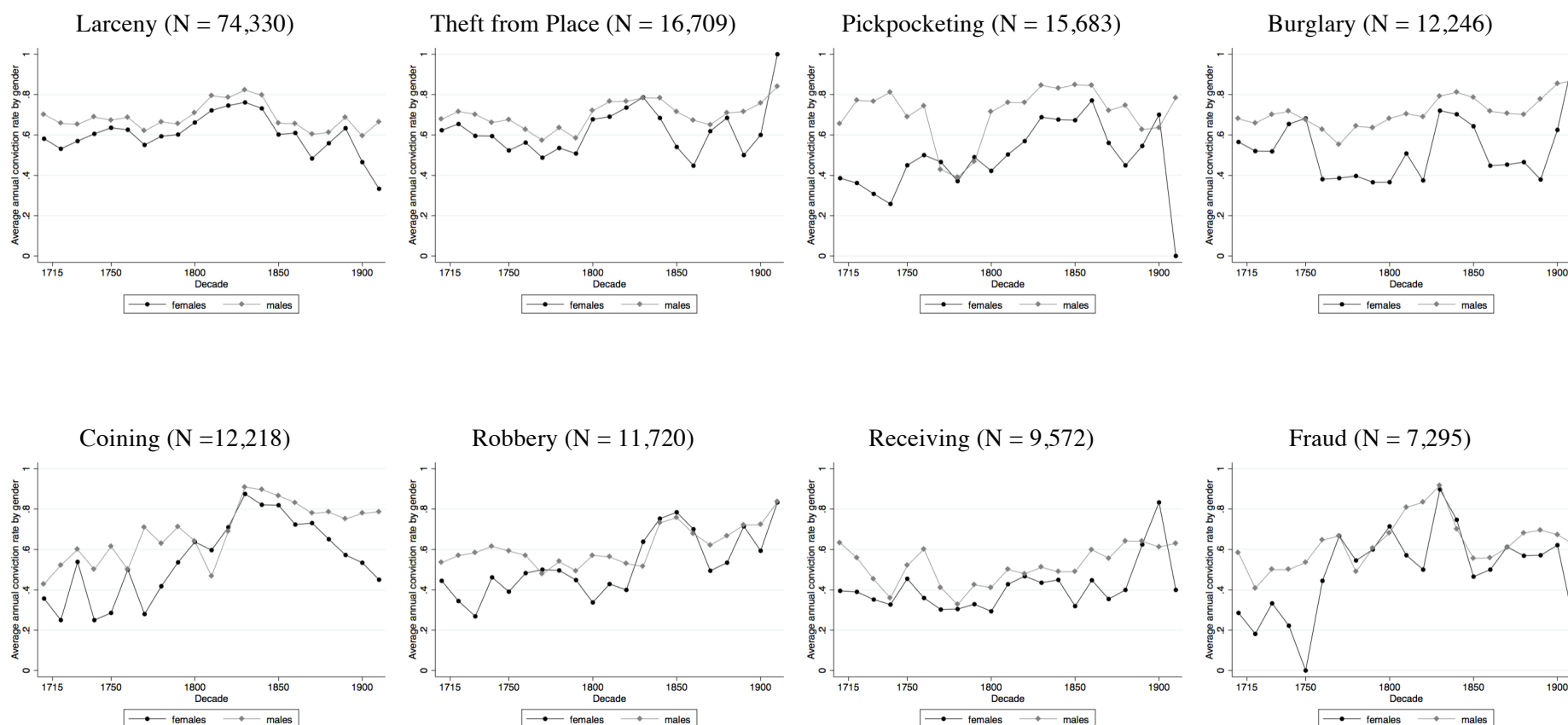


C. Raw gender gap in jury convictions + guilty pleas



NOTE – Figure A shows the average annual share of jury convictions of any offense by gender (averaged by decade), Figure B the average annual share of guilty pleas, and Figure C the combination of both jury convictions and guilty pleas. SOURCE - *The Old Bailey Proceedings Online* and own calculations.

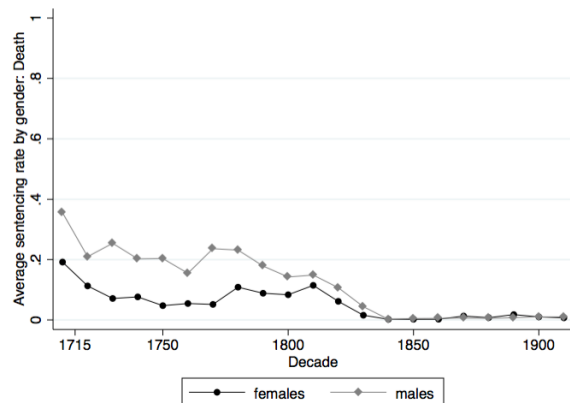
Figure 3. Raw Gender Gap in Jury Convictions for Most Common Offense Categories



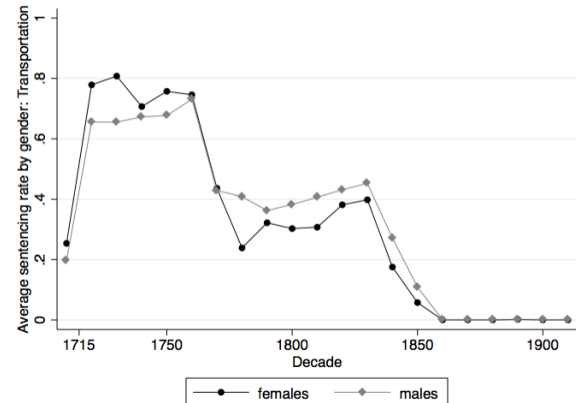
NOTE – Each graph shows the average annual share of jury convictions by gender (averaged by decade) for the eight most common offenses. Total sample sizes are noted above each figure. Note that although stealing from master was the 7th most common offense (N = 10,918), we do not present the figure here since the offense is only defined starting in the mid-1800s. SOURCE - *The Old Bailey Proceedings Online* and own calculations.

Figure 4. Raw Gender Gap in Sentencing

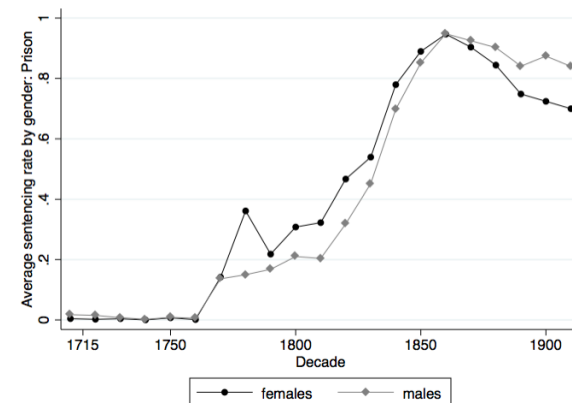
A. Capital punishment



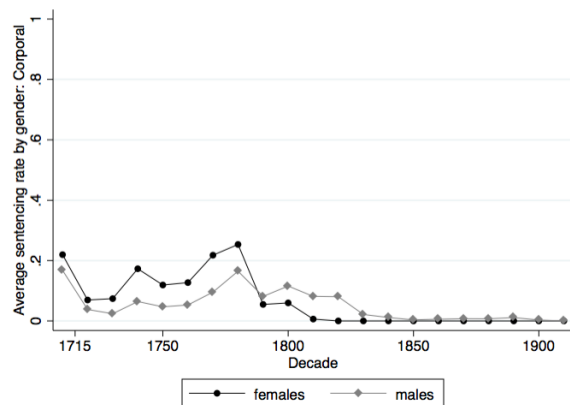
B. Transportation



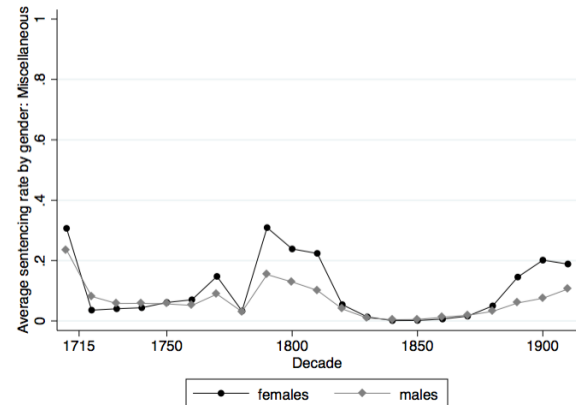
C. Prison



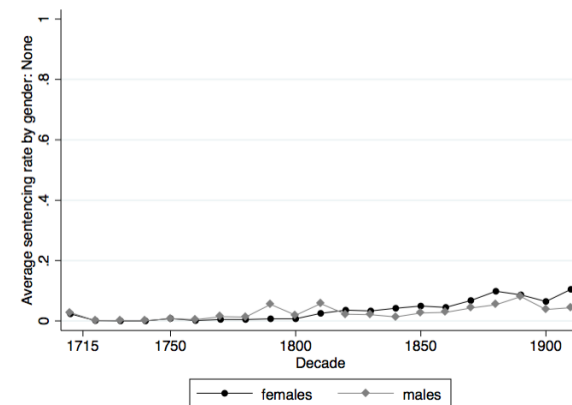
D. Corporal punishment



E. Miscellaneous



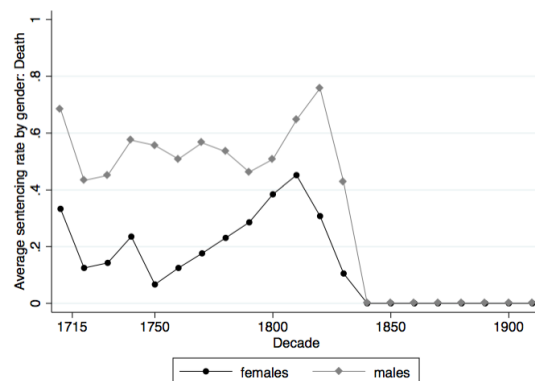
F. No punishment



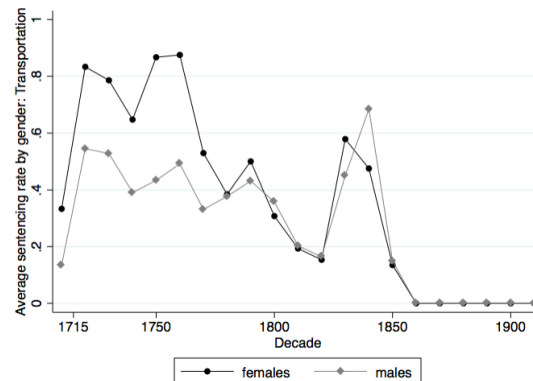
NOTE – The figures show the share of cases sentenced to capital punishment (A), transportation (B), prison (C), corporal punishment (D), miscellaneous punishment (E) and no punishment (F), each by gender. The share of cases is the annual share averaged by decade. SOURCE - *The Old Bailey Proceedings Online* and own calculations.

Figure 5. Raw Gender Gap in Sentencing for Burglary Offenses (Capital Offense until 1837)

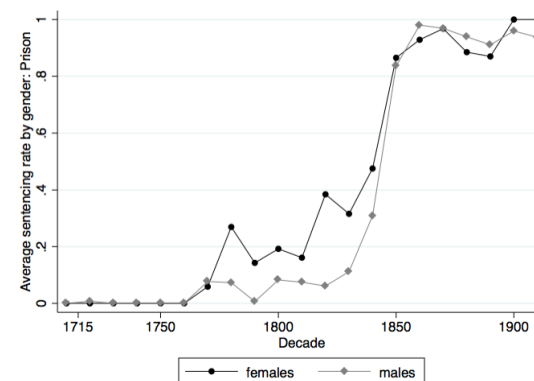
A. Capital punishment



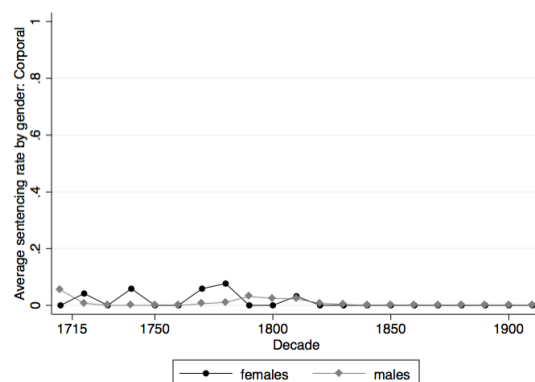
B. Transportation



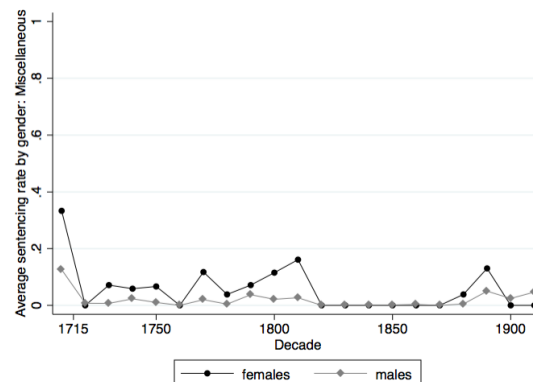
C. Prison



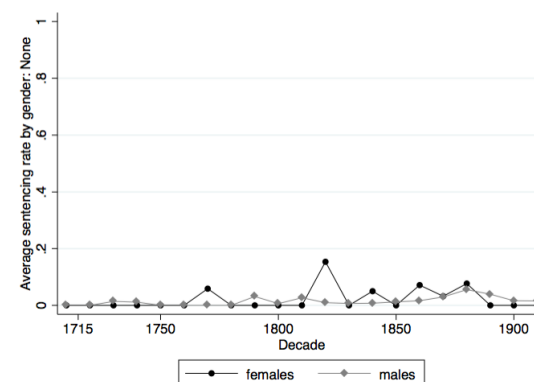
D. Corporal punishment



E. Miscellaneous



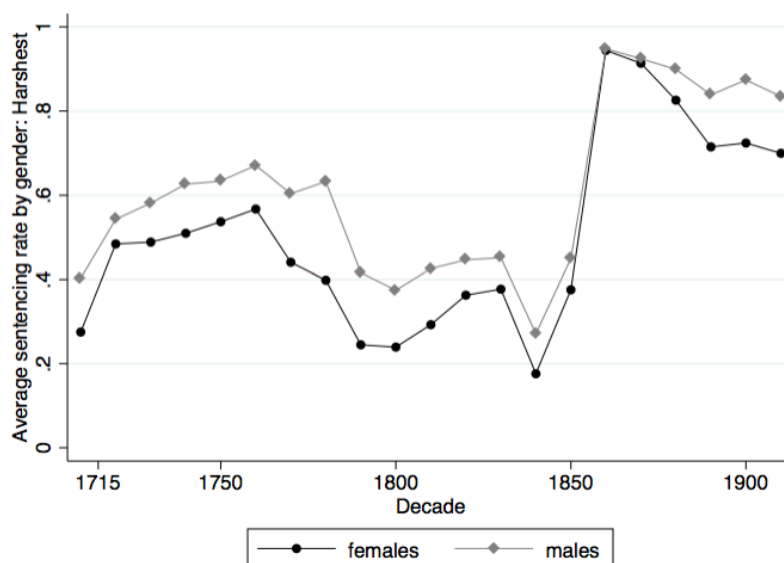
F. No punishment



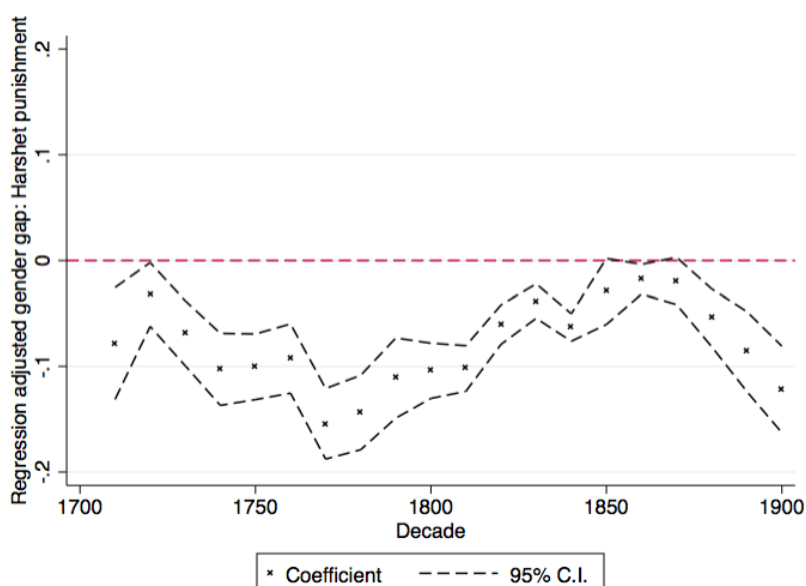
NOTE – The figures show the share of cases sentenced to capital punishment (A), transportation (B), prison (C), corporal punishment (D), miscellaneous punishment (E) and no punishment (F), each by gender, for all burglary convictions. The share of cases is the annual share averaged by decade. SOURCE - *The Old Bailey Proceedings Online* and own calculations.

Figure 6. Raw and Adjusted Gender Gaps in Receiving the Harshest Available Sentence

Panel A. Raw Gap in Harshest Sentence



Panel B. Adjusted Gap in Harshest Sentence



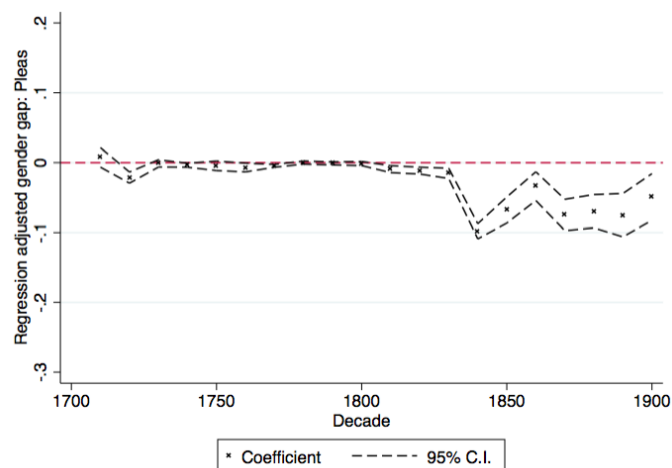
NOTE – These figures show the raw (panel A) and adjusted (panel B) gender gaps in the share of cases sentenced to the harshest punishment available for a given offense and a given point in time. In Panel B, each point represents the coefficient on the *female* dummy of a separate regression of the respective outcome on a set of controls (number defendants, offense, capital) for each decade. For capital offenses, the harshest punishment is the death penalty. For noncapital punishment, the harshest punishment is transportation when it is existent and imprisonment otherwise (i.e. during the American Revolution and after the abolition of transportation). The harshest punishment available is coded as missing during periods with uncertainty (aftermath of American Revolution and stage-wise abolition of transportation). SOURCE - *The Old Bailey Proceedings Online* and own calculations.

Figure 7. Regression Adjusted Gender Gap in Jury Convictions and Guilty Pleas

A. Adjusted gender gap in jury convictions



B. Adjusted gender gap in guilty pleas



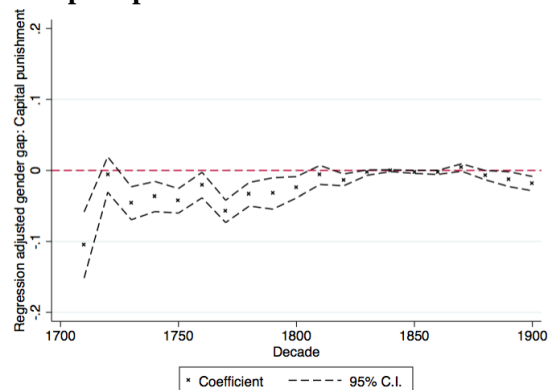
C. Adjusted gender gap in jury conviction and guilty plea



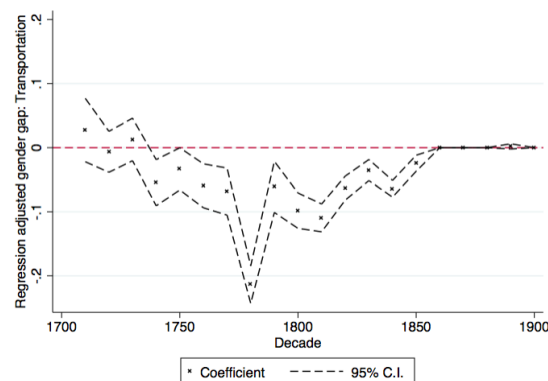
NOTE – The figures show the regression adjusted gender gaps in jury convictions (A), pleas (B) and jury convictions + pleas (C) by decade. Each point represents the coefficient on the *female* dummy of a separate regression of the respective outcome on a set of controls (number defendants, offense, capital eligibility) for each decade. SOURCE - *The Old Bailey Proceedings Online* and own calculations.

Figure 8. Regression Adjusted Gender Gap in Sentencing

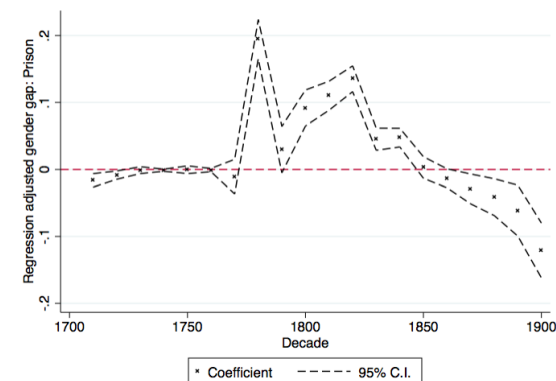
A. Capital punishment



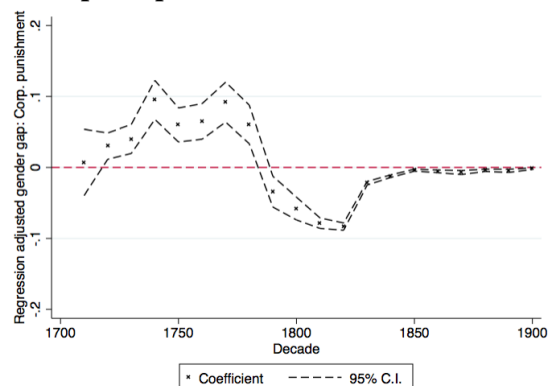
B. Transportation



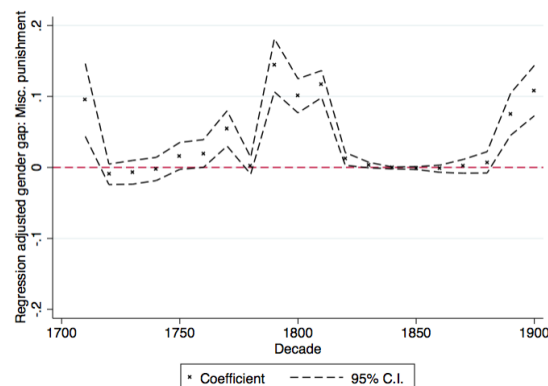
C. Prison



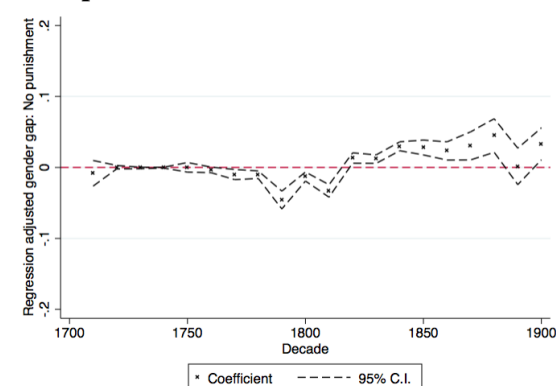
D. Corporal punishment



E. Miscellaneous



F. No punishment



NOTE – The figures show the regression adjusted gender gaps in sentences to capital punishment (A), transportation (B), prison (C), corporal punishment (D), miscellaneous punishment (E) and no punishment (F), each by gender. Each point represents the coefficient on the *female* dummy of a separate regression of the respective outcome on a set of controls (number defendants, offense, capital) for each decade. SOURCE - *The Old Bailey Proceedings Online* and own calculations.

Table 1. Summary Statistics

		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
		All	whole sample male	female	1715-1800 male	female	1801-1860 male	female	1861-1913 male	female
<u>Trial Characteristics</u>	N	202,569	157,226	44,634	31,643	14,751	83,027	24,232	42,556	5,651
capital	Mean	0.204	0.193	0.242	0.508	0.450	0.166	0.167	0.012	0.027
property off.	Mean	0.731	0.709	0.808	0.788	0.915	0.832	0.859	0.412	0.314
violent off.	Mean	0.114	0.126	0.072	0.141	0.048	0.066	0.045	0.230	0.253
sex off.	Mean	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
fraud off.	Mean	0.128	0.136	0.100	0.04	0.024	0.087	0.088	0.303	0.352
special off.	Mean	0.003	0.004	0.001	0.011	0.002	0.002	0.000	0.002	0.000
other off.	Mean	0.024	0.025	0.019	0.0198	0.012	0.013	0.008	0.054	0.082
plea	Mean	0.143	0.162	0.079	0.009	0.006	0.131	0.088	0.334	0.230
guilty (plea + guilty by jury)	Mean	0.726	0.748	0.650	0.629	0.548	0.776	0.708	0.784	0.668
<u>Verdict Conditional on Jury Trial</u>	N	173,502	131,810	41,118	31,346	14,669	72,114	22,097	28,350	4,352
guilty by jury	Mean	0.681	0.700	0.620	0.626	0.545	0.742	0.680	0.676	0.568
guilty by jury of initial charge	Mean	0.587	0.616	0.498	0.442	0.303	0.692	0.625	0.612	0.510
guilty by jury of lesser charge	Mean	0.098	0.087	0.130	0.191	0.259	0.051	0.059	0.064	0.059
acquitted by jury	Mean	0.318	0.299	0.379	0.374	0.455	0.258	0.320	0.320	0.430
guilty with recommendation for mercy	Mean	0.059	0.059	0.062	0.018	0.011	0.081	0.095	0.047	0.064
<u>Sentences Conditional on Conviction</u>	N	143,309	114,638	28,226	19,430	7,855	63,105	16,752	32,103	3,619
death penalty	Mean	0.066	0.071	0.047	0.217	0.085	0.058	0.037	0.007	0.009
transportation	Mean	0.285	0.277	0.319	0.521	0.555	0.343	0.278	0.000	0.000
prison	Mean	0.530	0.540	0.484	0.080	0.104	0.502	0.583	0.895	0.852
corporal punishment	Mean	0.040	0.039	0.045	0.090	0.151	0.040	0.006	0.007	0.000
harshest punishment	Mean	0.541	0.572	0.415	0.576	0.446	0.398	0.302	0.893	0.845

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. Gender Gaps in Jury Convictions and Guilty Pleas

sample	mean	(1) no controls	(2) + obs. controls	(3) + jury f.e.	(4) + judge f.e.
Panel A. Jury conviction of any charge					
1715 - 1900	0.681	-0.0790*** (0.0176)	-0.0803*** (0.0155)	-0.0824*** (0.0184)	-0.0818*** (0.0181)
<1751	0.607	-0.1118*** (0.0312)	-0.1234*** (0.0278)		
1751-1800	0.596	-0.0646*** (0.0142)	-0.0803*** (0.0127)	-0.0730*** (0.0174)	-0.0719*** (0.0170)
1801-1850	0.729	-0.0644*** (0.0199)	-0.0688*** (0.0192)	-0.0897*** (0.0232)	-0.0906*** (0.0229)
>1850	0.670	-0.0751*** (0.0240)	-0.0944*** (0.0131)		
Panel B. Jury conviction of initial charge					
1715 - 1900	0.587	-0.1170*** (0.0210)	-0.0979*** (0.0145)	-0.1034*** (0.0150)	-0.1026*** (0.0145)
<1751	0.279	-0.1415*** (0.0305)	-0.1168*** (0.0157)		
1751-1800	0.459	-0.1144*** (0.0184)	-0.1086*** (0.0084)	-0.1037*** (0.0117)	-0.1026*** (0.0109)
1801-1850	0.679	-0.0718*** (0.0234)	-0.0808*** (0.0217)	-0.0949*** (0.0236)	-0.0961*** (0.0236)
>1850	0.609	-0.0669** (0.0323)	-0.0939*** (0.0178)		
Panel C. Guilty plea					
1715 - 1900	0.129	-0.0727*** (0.0139)	-0.0361*** (0.0073)		
<1751	0.014	-0.0087*** (0.0020)	-0.0064** (0.0024)		
1751-1800	0.005	-0.0022*** (0.0008)	-0.0032** (0.0013)		
1801-1850	0.100	-0.0360*** (0.0051)	-0.0365*** (0.0060)		
>1850	0.296	-0.0918*** (0.0233)	-0.0596*** (0.0129)		
Observations Panel A and B					
1715 - 1900		167,255	167,255	55,023	54,711
<1751		15,618	15,618		
1751-1800		30,397	30,397	28,899	28,656
1801-1850		85,825	85,825	26,124	26,055
>1850		35,415	35,415		
Observations Panel C					
1715 - 1900		192,050	192,050		
<1751		15,838	15,838		
1751-1800		30,556	30,556		
1801-1850		95,318	95,318		
>1850		50,338	50,338		

NOTE – The table shows the estimated coefficients when regressing the outcome (any jury conviction – A, jury conviction of initial charge – B, guilty plea – C) on a dummy for female defendants by time period (separate regressions). Column (1) shows the results without additional controls, column (2) with observable controls (number defendants, capital eligibility, detailed offense category and year fixed effects), column (3) and (4) add jury and judge fixed effects for the subsample where the information is available (and no year fixed effects). The mean refers to the mean corresponding to the sample in column (1). Robust standard errors are clustered at the offense level and shown in parentheses next to the coefficient. *** p<0.01, ** p<0.05, *p<0.1

Table 3. Gender Gaps in Sentencing

		(1)	(2)	(3)	(4)	(5)	(6)	(7)
	mean (1)	no controls	+ obs. controls	+ jury & judge f.e.	+ obs. controls	+ jury & judge f.e.	+ obs. controls	+ jury & judge f.e.
		all	all	all	capital	capital	noncapital	noncapital
<i>Sentenced to harshest punishment available</i>								
<i>1715 - 1900</i>	0.523	-0.1447*** (0.0350)	-0.0747*** (0.0162)	-0.0980*** (0.0153)	-0.0781*** (0.0259)	-0.0705*** (0.0208)	-0.0648** (0.0233)	-0.1102*** (0.0044)
<i><1751</i>	0.528	-0.1006* (0.0553)	-0.0754*** (0.0214)		-0.0842* (0.0445)		-0.0663*** (0.0015)	
<i>1751-1800</i>	0.545	-0.1468*** (0.0390)	-0.1056*** (0.0122)	-0.1050*** (0.0112)	-0.1052*** (0.0279)	-0.0962*** (0.0253)	-0.1098*** (0.0011)	-0.1084*** (0.0003)
<i>1801-1850</i>	0.364	-0.0958*** (0.0190)	-0.0639** (0.0259)	-0.0924*** (0.0227)	-0.0462** (0.0172)	-0.0405* (0.0208)	-0.0621** (0.0288)	-0.1131*** (0.0078)
<i>>1850</i>	0.854	-0.0638* (0.0310)	-0.0420** (0.0160)		-0.1021*** (0.0143)		-0.0396** (0.0158)	
<i>Observations</i>								
<i>1715-1900</i>		129,450	129,450	34,478	24,373	13,623	105,077	20,855
<i>< 1751</i>		9,401	9,401		5,172		4,229	
<i>1751-1800</i>		16,654	16,654	16,056	7,545	7,150	9,109	8,906
<i>1800-1850</i>		70,629	70,629	18,422	11,131	6,473	59,498	11,949
<i>>1850</i>		32,766	32,766		525		32,241	

NOTE – The table shows the estimated coefficients when regressing the outcome on a dummy for female defendants by time period (separate regressions). The outcome variable is a dummy indicating whether the defendant is sentenced to the harshest punishment available. For capital offenses, the harshest punishment is the death penalty. For noncapital punishment, the harshest punishment is transportation when it is existent and imprisonment otherwise (i.e. during the American Revolution and after the abolition of transportation). The harshest punishment available is coded as missing during periods with uncertainty (aftermath of American Revolution and stage-wise abolition of transportation). Col. (4) and (5) are restricted to capital eligible cases, col. (6) and (7) to noncapital cases. Observable controls include number of defendants, detailed offense category, capital eligibility and year fixed effects. For further details, see Table 3. Robust standard errors are clustered at the offense level and shown in parentheses next to the coefficient. *** p<0.01, ** p<0.05, *p<0.1

Table 4. Blinder-Oaxaca Decomposition

	(1) 1715-1900 + year f.e.	(2) 1715-1900 + off f.e.	(3) 1715-1900 + ctrls	(4) <1751 + ctrls	(5) 1751-1800 + ctrls	(6) 1801-1850 + ctrls	(7) >1850 + ctrls	(8) >1850 + crim. hist.	(9) >1850 + age
Panel A. Jury conviction of any charge									
<u>Mean conviction rates</u>									
Males	0.7002***	0.7002***	0.7002***	0.6488***	0.6150***	0.7444***	0.6817***	0.6817***	
Females	0.6212***	0.6212***	0.6212***	0.5370***	0.5504***	0.6800***	0.6066***	0.6066***	
<u>Decomposition</u>									
Gap (males - females)	0.0790***	0.0790***	0.0790***	0.1118***	0.0646***	0.0644***	0.0751***	0.0751***	
Explained (Δ in characteristics)	0.0071	-0.0018	-0.0013	-0.0116	-0.0156	-0.0044	-0.0193	0.0015	
Unexplained (Δ in coefficients)	0.0719***	0.0808***	0.0803***	0.1234***	0.0803***	0.0688***	0.0944***	0.0736***	
<u>Decomposition in %</u>									
Gap (males - females)	100%	100%	100%	100%	100%	100%	100%	100%	
Explained (Δ in characteristics)	9%	2%	-2%	-10%	-24%	-7%	-26%	2%	
Unexplained (Δ in coefficients)	91%	98%	102%	110%	124%	107%	126%	98%	
N	167,255	167,255	167,255	15,618	30,397	85,825	35,415	35,415	
Panel B. Sentenced to harshest punishment available									
<u>Mean sentencing rates</u>									
Males	0.5524***	0.5524***	0.5524***	0.5609***	0.5830***	0.3840***	0.8617***	0.8617***	0.8726***
Females	0.4078***	0.4078***	0.4078***	0.4603***	0.4362***	0.2882***	0.7978***	0.7978***	0.8096***
<u>Decomposition</u>									
Gap (males - females)	0.1447***	0.1447***	0.1447***	0.1006*	0.1468***	0.0958***	0.0638**	0.0638**	0.0630**
Explained (Δ in characteristics)	0.0494**	0.0678**	0.0697*	0.0252	0.0400	0.0319*	0.0216	0.0243	0.0241
Unexplained (Δ in coefficients)	0.0953***	0.0769***	0.0750***	0.0754***	0.1068***	0.0639**	0.0422***	0.0395***	0.0389***
<u>Decomposition in %</u>									
Gap (males - females)	100%	100%	100%	100%	100%	100%	100%	100%	100%
Explained (Δ in characteristics)	34%	47%	48%	25%	37%	33%	34%	38%	38%
Unexplained (Δ in coefficients)	66%	53%	52%	75%	63%	67%	66%	62%	62%
N	129,450	129,450	129,450	9,401	16,654	70,629	32,766	32,766	30,853

NOTE – The table shows the results from a Blinder-Oaxaca decomposition (Stata command: oaxaca, pooled). Column (1) includes year dummies only, column (2) adds offense dummies, columns (3) to (7) add controls for capital eligibility and the number of defendants, column (8) adds criminal history and column (9) adds age (for Panel B only). *** p<0.01, ** p<0.05, *p<0.1

Table 5. Unobserved Differences in Case Characteristics

	(1) all cases no controls	(2) all cases + obs. controls	(3) all cases + wordcount controls	(4) all cases + wordcount group f.e.
<i>Panel A. Jury conviction of any charge</i>				
<i>1751-1810</i>	-0.0668*** (0.0168)	-0.0827*** (0.0191)	-0.0763*** (0.0171)	-0.0704*** (0.0149)
<i>Panel B. Jury conviction of initial charge</i>				
<i>1751-1810</i>	-0.1130*** (0.0186)	-0.1075*** (0.0085)	-0.1008*** (0.0069)	-0.0961*** (0.0062)
<i>Panel C. Sentenced to harshest punishment available</i>				
<i>1751-1810</i>	-0.1505*** (0.0316)	-0.1177*** (0.0151)	-0.1078*** (0.0170)	-0.0983*** (0.0189)
<i>Observations</i>				
Panel A and B	13,372	13,372	13,372	13,372
Panel C	7,697	7,697	7,697	7,697

NOTE – The table shows the estimated coefficients when regressing the respective outcome on a dummy for female defendants (separate regressions) using the sample including information on the word count of the trial. The sample includes the February, May and September sessions for the years 1751 to 1810. Column (1) includes no controls, column (2) observable controls (see Table 3 and 4), column (3) controls for the word count in addition to the observable controls and column (4) includes fixed effects for the quartile of the word count distribution instead of the linear word count control. Robust standard errors are clustered at the offense level and shown in parentheses next to the coefficient. *** p<0.01, ** p<0.05, *p<0.1

Table 6. Girlfriend Theory

	(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.	case f.e.
						same name	diff. name
<i>All offenses</i>							
All	-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)
N	111,892	55,363	44,295	11,068	11,068	1,998	4,364
<i>Property offenses</i>							
Burglary	-0.0439 (0.0268)	-0.3321*** (0.0233)	-0.2272*** (0.0490)	-0.2891*** (0.0325)	-0.2866*** (0.0304)	-0.2857** (0.1202)	-0.2150*** (0.0585)
N	4,631	4,821	4,035	786	786	63	230
housebreaking	-0.1133*** (0.0291)	-0.2234*** (0.0416)	-0.0803 (0.0697)	-0.2751*** (0.0609)	-0.2749*** (0.0570)		
N	2,032	1,630	1,370	260	260		
larceny_combined	-0.0517*** (0.0044)	-0.1244*** (0.0095)	-0.0960*** (0.0110)	-0.1047*** (0.0215)	-0.1063*** (0.0198)	-0.2000*** (0.0447)	-0.0419 (0.0285)
N	51,031	15,213	13,228	1,985	1,985	516	899
pickpocketing	-0.2032*** (0.0107)	-0.1628*** (0.0172)	-0.1998*** (0.0199)	0.0255 (0.0385)	0.0310 (0.0291)	-0.1463* (0.0833)	0.0400 (0.0343)
N	9,514	4,507	3,934	573	573	84	307
Receiving	-0.0808** (0.0363)	-0.1013*** (0.0127)	-0.0394* (0.0204)	-0.1758*** (0.0182)	-0.1775*** (0.0168)	-0.2042*** (0.0452)	-0.1470*** (0.0270)
N	1,114	7,396	4,469	2,927	2,927	411	1,092
stealingFromMaster	0.0572*** (0.0127)	-0.1371*** (0.0335)	-0.0986** (0.0458)	-0.1627** (0.0639)	-0.1453** (0.0716)	-0.2105 (0.1379)	-0.0857 (0.0999)
N	5,559	1,763	1,522	241	241	91	95
theftFromPlace	-0.0647*** (0.0101)	-0.1299*** (0.0191)	-0.1157*** (0.0253)	-0.0753** (0.0331)	-0.0787** (0.0311)	-0.1982*** (0.0655)	0.0117 (0.0461)
N	10,202	4,426	3,603	823	823	225	353
<i>Violent offences</i>							
robbery_combined	-0.1256*** (0.0256)	-0.0272 (0.0175)	-0.0237 (0.0233)	-0.0140 (0.0300)	-0.0032 (0.0186)	-0.1081** (0.0521)	0.0060 (0.0289)
N	4,641	5,974	5,114	860	860	75	342
<i>Fraud offences</i>							
coiningOffences	-0.0552*** (0.0124)	-0.1022*** (0.0169)	-0.0743*** (0.0269)	-0.1306*** (0.0238)	-0.1058*** (0.0230)	-0.2976*** (0.0688)	-0.0040 (0.0297)
N	5,373	3,340	1,949	1,391	1,391	201	566
forgery	-0.0840** (0.0391)	-0.0686 (0.0441)	-0.0012 (0.0951)	-0.0063 (0.0508)	0.0089 (0.0424)		
N	2,652	961	735	226	226		
fraud	0.0028 (0.0313)	-0.1407*** (0.0456)	0.0388 (0.1061)	-0.1065* (0.0572)	-0.0885* (0.0451)		
N	2,585	1,538	1,306	232	232		

NOTE – The table shows estimated coefficients when regressing jury conviction for any charge on a dummy for female. Col.(1) is restricted to single defendant, (2)–(5) to multiple defendant and (6)–(7) to two defendant cases. Col.(3) is restricted to same gender cases, (4)–(7) to mixed gender cases. Col.(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 for the pooled regression (robust for offense specific regression) and shown in parentheses. *** p<0.01, ** p<0.05, *p<0.1

Table 7. Share of Female Defendants over Time and Across Sessions

	(1) Decreasing share female + obs. controls	(2) Increasing share female + obs. controls	(3) High session share female + obs. controls	(4) Low session share female + obs. controls
Panel A. Jury conviction of any charge				
1715-1750	-0.1231*** (0.0084)	-0.2400*** (0.0629)	-0.1230*** (0.0113)	-0.1241*** (0.0121)
1751-1800	-0.0807*** (0.0064)	-0.1159 (0.0733)	-0.0771*** (0.0085)	-0.0823*** (0.0097)
1801-1850	-0.0701*** (0.0037)	-0.0734*** (0.0259)	-0.0652*** (0.0048)	-0.0724*** (0.0055)
>1850	-0.0993*** (0.0082)	-0.0780*** (0.0172)	-0.0906*** (0.0088)	-0.1036*** (0.0120)
Observations				
1715-1750	14,507	508	7,777	7,841
1751-1800	28,677	571	15,078	15,319
1801-1850	81,090	2,175	43,101	42,724
>1850	26,476	4,925	17,821	17,594
Panel B. Sentenced to harshest punishment available				
1715-1750	-0.0768*** (0.0093)	0.1267 (0.1320)	-0.0699*** (0.0126)	-0.0805*** (0.0135)
1751-1800	-0.1081*** (0.0080)	0.0396 (0.1182)	-0.1117*** (0.0104)	-0.1044*** (0.0123)
1801-1850	-0.0623*** (0.0041)	-0.0446* (0.0270)	-0.0669*** (0.0054)	-0.0612*** (0.0060)
>1850	-0.0275*** (0.0053)	-0.0505*** (0.0169)	-0.0325*** (0.0060)	-0.0572*** (0.0096)
Observations				
1715-1750	8,859	253	4,599	4,802
1751-1800	15,837	275	8,292	8,362
1801-1850	67,642	1,221	34,951	35,678
>1850	25,012	3,183	14,800	17,966

NOTE – The table shows the estimated coefficients when regressing the outcome (jury conviction for any charge in Panel A and sentenced to harshest punishment available in Panel B) on a dummy for female defendants. All columns include observable controls as in Table 3. Column (1) includes offenses with a decreasing share of female defendants over time (burglary, coining, fraud, housebreaking, larceny, pickpocketing, receiving, robbery, shoplifting, stealing from master, theft from place); column (2) includes offenses with an increasing share of female defendants over time (manslaughter, murder, wounding). Columns (3) and (5) split the sample for each year group depending on whether the session had a share of female defendants that is above or below the median within that year group. Robust standard errors are clustered at the offense level and shown in parentheses. *** p<0.01, ** p<0.05, *p<0.1

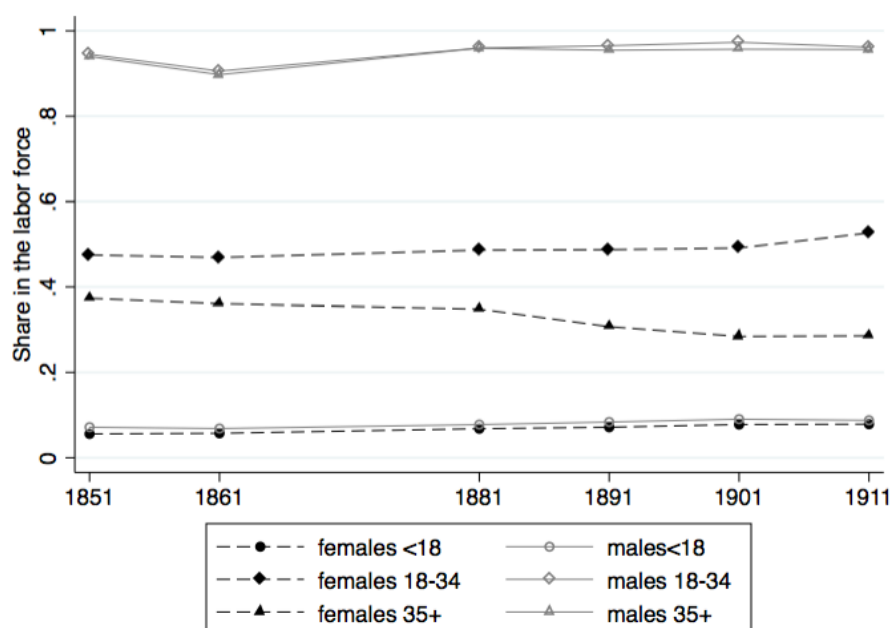
Table 8. Offenses with High and Low Shares of Female Defendants

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

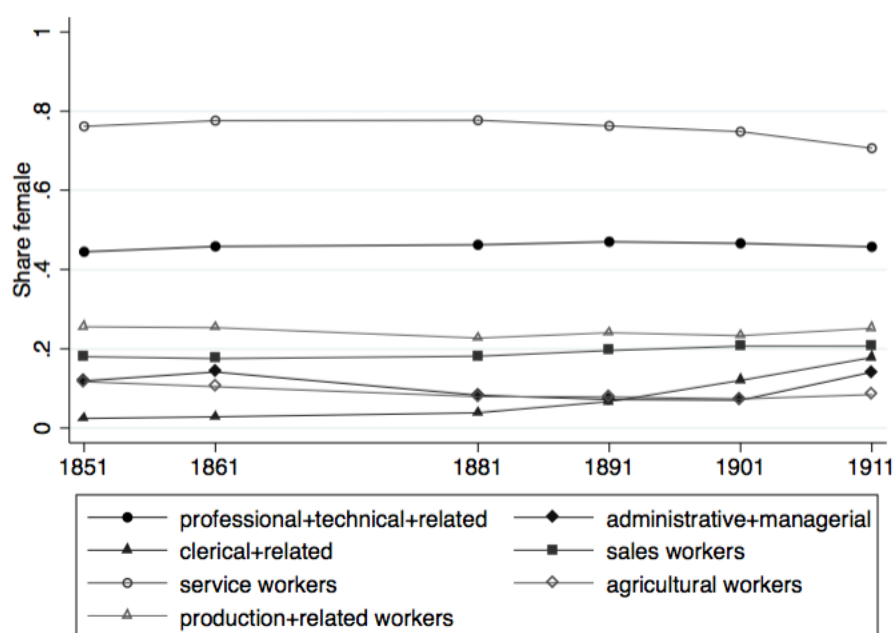
NOTE – The table shows the estimated coefficients when regressing the outcome (jury conviction for any charge) on a dummy for female defendants, by detailed offense. Column (1) lists the offense; column (2) shows the number of observations (1715-1900), column (3) the share females, column (4) the estimated coefficient and column (5) the standard error. 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

Appendix Figure 1. Female Participation in Labor Markets 1851 - 1911

A. Labor force participation by age and gender



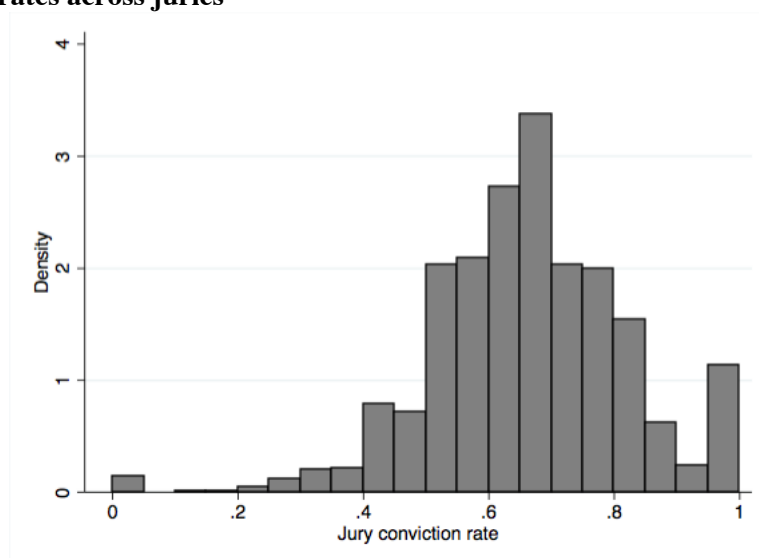
B. Share females by occupational group



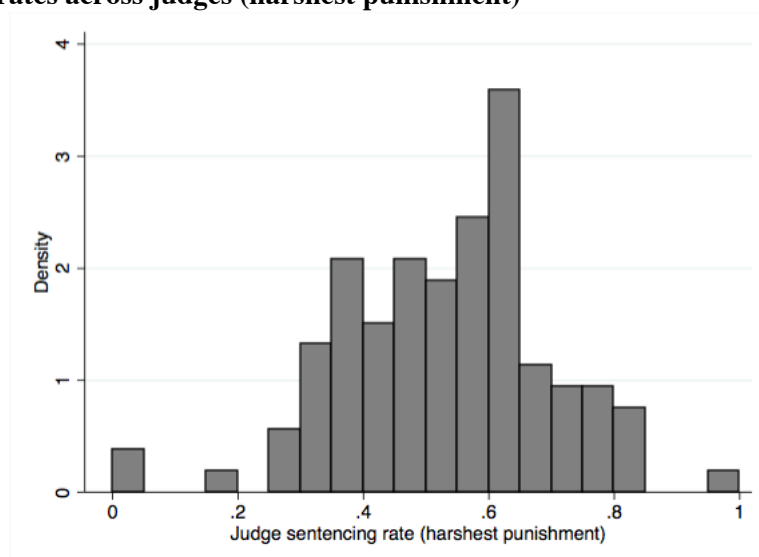
NOTE – The figures show the share of participants in the labor force by gender and age group (A) and the share of females by occupational group (B). All figures are based on the London and Middlesex population, for the census years 1851, 1861, 1881, 1891, 1901 and 1911.

Appendix Figure 2. Variation Across Juries and Judges

Panel A. Conviction rates across juries



Panel B. Sentencing rates across judges (harshest punishment)



NOTE – The figure shows histograms of the within jury conviction rates (Panel A) and the within judge sentencing rate to the harshest punishment available (Panel B). Both histograms are based on the reduced sample for which jury and judge identifiers are available. SOURCE - *The Old Bailey Proceedings Online* and own calculations.

Appendix Table 1. Occupational Differences Across Gender

1851: Occupation	N	Share females	1901: Occupation	N	Share females
Total (w/o known occupation)	2 368 643	55%	Total (w/o known occupation)	5 135 793	54%
Occupation known	1 101 217	37%	Occupation known	2 407 390	34%
No occupation/unknown	1 267 426	70%	No occupation/unknown	2 728 403	71%
<u>20 most common occupations</u>					
House servants nfs and maids	163 054	87%	House servants nfs and maids	242 155	93%
Dealer, merchant etc. (wholesale and retail trade)	75 206	18%	Dealer, merchant etc. (wholesale and retail trade)	186 979	25%
Common labourers or general labourers	51 778	3%	Office clerks, specialisation unknown	107 982	18%
Boot and shoe makers and repairers	38 192	24%	Common labourers or general labourers	79 342	1%
Washing and laundry services	37 547	98%	Dressmakers	69 841	99%
Messengers	34 955	2%	Delivery men and drivers of goods	67 897	0%
Dressmakers	33 022	99%	Washing and laundry services	60 667	93%
Tailors and tailoresses	28 838	29%	Messengers	54 212	3%
Carpenter and joiner	23 144	2%	Cook (domestic) UK only	52 163	97%
Others hand sewers	20 397	99%	Tailors and tailoresses	48 779	54%
Delivery men and drivers of goods	18 744	2%	Painters, construction	38 896	1%
Office clerks, specialisation unknown	18 067	1%	Carpenter and joiner	38 057	0%
Weavers	15 613	55%	Boot and shoe makers and repairers	33 469	19%
Bricklayers	13 901	1%	Bricklayers	33 118	0%
Cook (domestic) UK only	13 539	95%	Charworkers	28 094	100%
Milliners	11 925	97%	Pipe fitters	25 292	0%
Farm workers, specialisation unknown	11 821	7%	Other freight handlers	25 059	3%
Painters, construction	11 800	5%	Teachers not in higher education	24 869	72%
Charworkers	10 891	99%	Others hand sewers	23 686	94%
<i>Share 20 most common of known occupations</i>		57%	<i>Share 20 most common of known occupations</i>		52%

NOTE – The table shows the number of individuals in the London and Middlesex population with/without known occupation as reported in the 1851 and 1901 Census, as well as the number of individuals employed in the 20 most common occupations and the share females in that occupation. The last row shows the share of employees in the 20 most common out of all known occupations. Milliner: Person who makes or sells women's hats; charworkers: cleaning woman.

Appendix Table 2. Sample Creation

Detailed offence category (combined)	N	Share females	N	Share females	N	Share females	N	Share females	Drop
	1715-1750		1751-1800		1801-1850		1851-1913		
Against crown (major)	18	0.33	11	0.45	26	0.00	27	0.00	D1
Against crown (minor)	81	0.07	105	0.01	77	0.00	18	0.11	D1
Animal theft	535	0.03	841	0.02	2141	0.02	930	0.01	D2
Arson	16	0.31	27	0.37	68	0.31	456	0.22	
Assault	31	0.16	6	0.17	415	0.06	656	0.06	
Bigamy	165	0.29	89	0.18	446	0.16	1508	0.17	
Breaking into place	0	-	0	-	0	-	1	0.00	D1
Burglary	892	0.15	2229	0.09	3063	0.08	6062	0.05	
Coining offences	95	0.46	569	0.32	3054	0.31	8500	0.26	
Embezzlement	5	0.00	14	0.00	2953	0.03	1595	0.01	D2
Forgery	170	0.21	406	0.07	1438	0.09	4476	0.08	
Fraud	105	0.32	286	0.11	1188	0.13	5716	0.07	
Habitual criminal	0	-	0	-	0	-	14	0.07	D1
Housebreaking	393	0.28	323	0.21	2433	0.12	1346	0.07	
Infanticide (combined)	77	0.97	50	0.96	141	0.94	769	0.87	D2
Larceny (combined)	6734	0.42	16323	0.32	43816	0.24	7457	0.14	
Libel	3	0.00	3	0.00	67	0.07	579	0.13	
Mail	0	-	14	0.00	170	0.01	1397	0.02	D2
Manslaughter	13	0.08	26	0.04	494	0.11	1632	0.21	
Missing	67	0.25	45	0.18	40	0.20	144	0.13	D3
Murder	511	0.12	473	0.12	484	0.19	895	0.25	
Other (combined)	2249	0.49	601	0.31	1352	0.18	2610	0.13	D3
Perjury	146	0.24	234	0.20	212	0.15	708	0.20	
Perverting justice	37	0.14	132	0.21	187	0.19	278	0.10	
Pickpocketing	1022	0.58	796	0.54	11042	0.24	2823	0.17	
Rape	106	0.04	141	0.06	300	0.01	1453	0.01	D2
Receiving	675	0.48	1831	0.36	5581	0.28	1485	0.12	
Return from transportation	57	0.12	168	0.05	127	0.03	19	0.00	
Riot	38	0.08	106	0.05	34	0.06	84	0.02	
Robbery (combined)	1556	0.17	2503	0.13	2474	0.18	5187	0.09	
Sexual assault (combined)	8	0.00	0	-	272	0.00	958	0.00	D2
Shoplifting	981	0.39	831	0.56	1601	0.45	100	0.22	
Sodomy (combined)	53	0.00	31	0.00	213	0.00	1082	0.00	D2
Stealing from master	0	-	0	-	7715	0.27	3203	0.14	
Theft from place	2379	0.45	3146	0.35	8294	0.22	2890	0.07	
Wounding	7	0.00	74	0.01	1246	0.19	4423	0.14	

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.